DEVELOPMENT AND TESTING OF COMPUTATIONAL
PROCEDURES FOR SIGNAL TIMING DESIGN AT ISOLATED
INTERSECTIONS

By

JIN-TAE KIM

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2001

To my beloved parents and wife

ACKNOWLEDGMENTS

I would like to take this opportunity to thank my advisor, Professor Kenneth G. Courage, for his continual advice, guidance, support and, moreover, for honoring me by chairing my supervisory committee during the entire course of my doctoral research. I would like to express my extreme appreciation not only for his technical guidance but for the warmth I felt under his leadership.

I would like to express my gratitude to the rest of my supervisory committee, Dr. Gary Long, Dr. Scott Washburn, Dr. Joseph P. Geunes and Dr. Anne M. Wyatt-Brown, for their encouragement, comments, and interest in my research.

I would like to extend my appreciation to Dr. Charles E. Wallace, the director of the Transportation Research Center. Although he was not on my committee, I would like to thank him for his encouragement and concern during the course of my doctoral study.

I also would like to thank Phil Hill, who helped me while I was working with the Center for Microcomputers in Transportation (McTrans). His knowledge in computer technology inspired me during the research.

My special deep gratitude goes to Professor Myoung-Soon Chang, who taught and guided me when I was an undergraduate student at Hanyang University, Korea, for his encouragement and support. He suggested that I study abroad and helped me to develop my keen interest in research.

I also want to express my deep appreciation to my parents, Hoon-Ho Kim and Soon-Rea Cho, for their patience and for supporting me in many ways with their love.

I would like to thank my wife, Hee-Jae Na, who has been willing to share the good and bad moments during my doctoral study. Her patience and love have provided me with the warmest feeling and the best support to complete my research.

Finally, I would like to thank God. Without his love and concern, I would not have this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

xii

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

DEVELOPMENT AND TESTING OF COMPUTATIONAL
PROCEDURES FOR SIGNAL TIMING DESIGN AT ISOLATED
INTERSECTIONS

By

Jin-Tae Kim

December, 2001

Chairman: Kenneth G. Courage
Major Department: Department of Civil Engineering

Traffic control signals are an important element of urban street systems. While
they are not, in themselves, the cause of delay and other problems perceived by the
motorist, their operation has a very strong influence on the performance of these systems.
There are several elements of the signal operation that may be addressed to improve the
performance at any given location. One of the key elements is the accuracy of
performance estimators themselves. Another is the manner in which the performance
estimators are used to achieve the "best" signal-timing plan. Both of these elements are
addressed in this dissertation.

The Highway Capacity Manual (HCM) describes a procedure for evaluating the
performance of a signal in terms of control delay, defined as the amount of delay that
would be eliminated for a given traffic movement if the signal control were eliminated
from that movement. The HCM signalized intersection model represents the state of

current knowledge of traffic signal operation. This dissertation explores the potential for advancement by addressing three specific shortcomings of the HCM model and its application: (1) the accuracy of the estimation procedure for the average green times for traffic actuated control, (2) the need for an iterative technique that can achieve the "best" performance of an intersection as evaluated by the HCM procedure and (3) potential improvements to the HCM model structure through the use of a "time-scan" technique to represent queue accumulation and departure. The following activities are described in the dissertation:

1. The functional and computational requirements for a formal optimization model were proposed, and an algorithm to meet those requirements was developed and tested. The optimization procedure used a hybrid genetic algorithm, involving a combination of a pure genetic algorithm and a "hill-climbing" approach.

2. The HCM procedure for average phase-length estimation was modified to improve its performance. Comparisons were made between the signal timings designed by the existing and proposed methods.

3. A time-scan procedure for representing the accumulation and discharge of queues on an approach to a signalized intersection was developed and tested using a hypothetical example involving permitted left turns.

The proposed enhancements were evaluated using the CORSIM simulation model as a surrogate for field data collection. The results indicated that the proposed optimization scheme was able to achieve its design objectives, the green time estimation procedures were improved by the proposed enhancements, and the accuracy of the delay

estimates for approaches involving permitted left turns was improved by the proposed time-scan model structure.

CHAPTER 1
INTRODUCTION

Traffic control signals are an important element of urban street systems. While they are not, in themselves, the cause of delay and other problems perceived by the motorist, their operation has a very strong influence on the performance of these systems. It is not surprising, then, that substantial efforts and resources have been focused on maximizing the effectiveness of a number of traffic signals in the United States.

There are several elements of the signal operation that may be addressed to improve the performance at any given location. One of the key elements is the accuracy of performance estimators themselves. Another is the manner in which the performance estimators are used to achieve the "best" signal-timing plan. Both of these elements will be addressed in this dissertation.

The Highway Capacity Manual (HCM) [1] describes a procedure for evaluating the performance of a signal in terms of control delay, which is measured in seconds per vehicle and defined as the amount of delay that would be eliminated for a given traffic movement if the signal control were eliminated from that movement. Again, that is not to say that the signal has caused the delay, because delay is inevitable when traffic movements compete for right of way. For the same reason, the signal is not able to eliminate all delay from the intersection.

The signal operation has two possible objectives with respect to delay: It could try to minimize the aggregate delay to all vehicles using the intersection, or it could try to

1

distribute the delay equitably among competing movements. Both of these objectives are legitimate and have been employed to varying degrees in contemporary traffic control systems. Both objectives will be addressed in this dissertation.

1.1 Problem Statement

The HCM procedure was developed with Transportation Research Board (TRB) oversight. It embodies the consensus of national and international experts in the development of the performance criteria and estimation procedures. It is therefore widely accepted as the standard for performance estimation on all types of transportation facilities, including signalized intersections [2, 3].

The HCM signalized intersection model represents the state of current knowledge of traffic signal operation. The document is updated periodically to reflect advances in the area. This dissertation will explore the potential for advancement by addressing three apparent shortcomings of the model and its application.

1.   The HCM procedure for signalized intersections requires a knowledge of two signal timing parameters:

   a.   the cycle length, defined as the time required to execute a complete sequence of displays to accommodate all movements; and

   b.   the proportion of effective green time available to each movement. If the signal operates on a "pretimed" plan (i.e., all green times are specified), the required parameters may be obtained from the timing plan. The vast majority of signals today, however, operate on the "traffic-actuated" principle, which means that the amount of green time displayed to each movement will depend on information from traffic detectors. Therefore, with traffic-actuated control, the two required signal timing

parameters will be undefined and must therefore be estimated by another procedure. There is room for improvement in the current estimation procedure prescribed in the HCM.

2.   Because the timing plan must be known before delay can be estimated, the design of an optimal timing plan requires an iterative procedure. The HCM does not suggest a procedure, and no such procedure currently appears in the literature. The performance estimation procedures used in all of the documented optimization models deviates from the HCM procedure. Software products that purport to use the HCM procedure are based on undocumented and informal "trial and error" techniques. The result is that many signal timing plan designs are not optimal with respect to the standard that has been adopted by many agencies. A formal optimization procedure that is built around the HCM performance estimation technique would represent a definite advance in the state of the art.

3.   Estimation of delay involves the analysis of the queue accumulation and service over a complete cycle of operation. The HCM model uses a deterministic approximation known as the "queue accumulation polygon", a geometric shape whose area represents the control delay. More advanced models have adopted a "time-scan" approach in which the queue accumulation and service characteristics are updated at the end of each second of operation. The time-scan procedure is intuitively more accurate and therefore deserves exploration.

1.2 Objectives

The overall goal of this study was to address the three problem areas identified above by developing and testing proposed concepts and algorithms that show some promise for improvement. Specifically, the objectives were as follows:

1. Identify specific shortcomings of the existing HCM procedure for estimating the average cycle length and green times for traffic-actuated control as a function of traffic volumes, intersection characteristics and controller settings. Propose and evaluate concepts and algorithms that could improve this procedure.

2. Develop the concepts and algorithms required to produce a formal optimization technique based on the HCM delay estimation procedure.

3. Evaluate the timing plans produced by this technique with respect to existing models and software products.

4. Explore the potential improvement in the accuracy of the current HCM model through the substitution of a time-scan-based procedure for representing the accumulation and discharge of queues on an approach to a signalized intersection.

1.3 Organization

The following tasks were carried out in support of these objectives and the results are reported in their respective chapters:

1. A literature review was performed to describe the current body of knowledge in signal-timing design for both pretimed and actuated operations and heuristic search methods and to acquire the necessary insight to advance the current state of knowledge. The results of this task

are presented in Chapter 2, along with an explanation of the basic operation of traffic-actuated controllers.

2. The functional and computational requirements for a formal optimization model were proposed, and an algorithm to meet those requirements was developed and tested, as discussed in Chapter 3.

3. The HCM procedure for average phase-length estimation was modified to improve its performance. Based on the modified estimation procedure, the relationship between the maximum green time parameter, a key determinant of the signal timing, and the operation of the intersection was verified. The results of this task are presented in Chapter 4.

4. Performance comparisons were made between the signal timings designed using existing and proposed methods. The green times and the maximum green times for both pretimed and traffic-actuated operation were addressed in these comparisons, which are presented in Chapter 5.

5. A time-scan-based procedure for representing the accumulation and discharge of queues on an approach to a signalized intersection was proposed to improve the accuracy of the current HCM model, especially in the evaluation of permitted left turns. An assessment of the potential of the proposed procedure for improvement of the HCM modeling process is presented in Chapter 6.

6. The findings and conclusions of the study are summarized and presented in Chapter 7, along with recommendations for future research.

This dissertation attempts to advance the state of knowledge on the subject of signal timing optimization and performance evaluation. Because the subject matter is somewhat complex, some tutorial information is included in the literature review presented in Chapter 2. With the exception of this material, the remainder of the discussion presented herein assumes that the reader has a basic knowledge of the principles of traffic engineering in general and of traffic control in particular.

CHAPTER 2
LITERATURE REVIEW

This chapter presents a review of the literature and identifies the underlying background and knowledge on topics related to this research study. It focuses generally on previous research related to the topic of the dissertation and specifically on the background required to understand the proposed computational procedures. It emphasizes the research on signal-timing design and performance analysis of isolated signalized intersections and discusses recent progress that has been made in these areas.

The material is organized into six topics including (1) an explanation of how modern traffic-actuated signal controllers operate, (2) current signal-timing-optimization strategies, (3) current procedures for estimating the average green time for a traffic-actuated phase as a function of the traffic volume and controller settings, (4) principles and procedures for optimizing the maximum green time setting for a traffic-actuated phase, (5) current computer-based models and software used for designing and evaluating traffic signal timing plans and (6) search techniques that may be employed to promote optimization of traffic control parameters.

Because of the complexity of the topic, the literature review also includes some background required to understand (1) the mechanics of traffic-actuated operation, (2) heuristic searching methods and (3) existing computer-based signal-timing-design models.

2.1 Traffic-Actuated Operation

Traffic-actuated phases fundamentally can be operated with three basic parameters: minimum green time, maximum green time and unit extension time [4, 5]. Additional control parameters are required for the "volume-density" mode of operation for dynamic changes in the minimum green time and unit extension time for each cycle [4, 5]. The minimum green time is the initial length of green time displayed when the phase is called. A phase can be skipped if no vehicle actuation is received from the detector installed on the subject approach in order to provide the right-of-way to the next phase. If vehicle actuations occur continuously after the minimum green time is displayed, the green time is extended based on a combination of vehicle actuations and the unit extension time. The unit extension time is the time period that a vehicle actuation should have occurred to add one more unit extension time to the green. If not, the phase terminates. If an actuation occurs during a unit extension period, another unit extension is provided to the green from the time that actuation is detected. This step can continue until the maximum green time is reached. If no vehicle for a unit extension is detected before the maximum green time, the gapout occurs – the green terminates before it reaches the maximum green time. The maximum green time is the upper limit of the green time extension. Figure 2-1 represents the basic mechanics of traffic-actuated operation.

In the volume-density mode, which is effective when neighboring intersections are relatively far enough away to yield dispersion of vehicle platoons, usually with high approach speeds, the minimum green time and the unit extension time are dynamically determined based on the actuations [4, 5]. This requires such additional parameters as
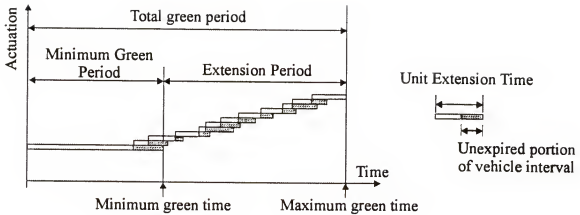
Figure 2-1: Schematic of actuated phase interval

initial minimum green, seconds per actuation, time before reduction and time to reduce. Several operational strategies exist, and the values of those parameters should be specified based on the strategy. For example, the initial minimum green can be determined for three different types of operations: extensible, added and computed initial types [4, 6].

In addition, a traffic-actuated controller requires another group of control parameters that are required for the actuated operation in a dual-ring scheme. They include minimum/maximum recall, pedestrian recall and detector memory [4, 6]. Traffic actuated control with dual-ring phase operation requires all such input parameters described above. Designing all those parameters currently requires a trial-and-error searching method over a whole set of feasible parameters based on simulation analysis, since no analytical model is available for sensitivity study in those parameters.

The phases of a traffic controller can be represented and operated by either a single-ring scheme or a dual-ring-phase scheme. A dual-ring-phase scheme provides wide flexibility in terms of the operation of both pretimed or actuated control, while a single-ring scheme is limited with the traffic-actuated operation [4, 5, 6]. Traffic-actuated operation has been utilized mostly with the dual-ring scheme due to its functionality. The dual-ring-phase scheme, composed of eight phases, is illustrated in Figure 2-2. Each of these phases is assigned to either left-turn or through movement in the approaches to an intersection. The standard numbering and movement assignments for those eight phases are illustrated in Figure 2-2.
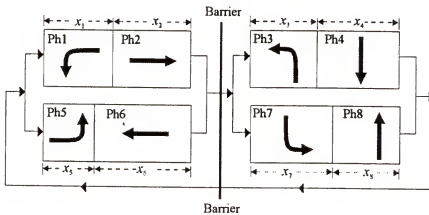
Figure 2-2: Dual-ring phase diagram

The left side of dual ring represents the beginning of a cycle, and the right side represents the ending of the cycle. The small arrows in the figure represent the time flowing over a cycle. The bold arrows represent the movements assigned to each phase. The variable x_i represents the length of the phase i. The east-westbound movements are assigned to the left side of the barrier, and the north-southbound movements are assigned to the right side of the barrier as a standard. A barrier is set in between east-westbound

and north-southbound movements. The phases must be simultaneously terminated to cross the barrier.

When a cycle begins, the eastbound and westbound left-turn movements, phases 1 and 5, possess the right-of-way (phases 1 and 5 display the green). When either of those phases terminates due to either gap out (no actuation detected for a period of unit extension time) or the maximum green time, the adjacent phase starts receiving the right-of-way. For example, when phase 1 terminates first, phase 2 turns to green. In such a case, the eastbound through movement, phase 2, flows with the eastbound left-turn movement, phase 5.

The sum of the lengths of either the phases numbered from 1 to 4 or the phases numbered from 5 to 8 represents the cycle length. Those two rings cycle by synchronizing at the two points of time: the barrier and the beginning of a cycle. Those two rings are a feature of the dual-ring phase operation.

2.2 Signal Timing Optimization Strategies

Several design strategies can be applied depending on traffic conditions when traffic signal timing parameters are designed for pretimed operation. Such strategies include (1) equalizing degree of saturation for critical lane groups, (2) minimizing aggregated delay of all vehicles and (3) equalizing the Level of Service (LOS) of critical lane groups [1, 7]. The following subsections present the review of those signal timing optimization strategies.

2.2.1 Equalize Degree of Saturation for Critical Lane Groups

The strategy of equalizing degree of saturation for critical lane groups is designed primarily for pretimed operation [1, 7]. The objective of the strategy is to balance the operational conditions of critical lane groups by assigning green time to phases, based on

a target volume-to-capacity (v/c) ratio. In short, the strategy is to equalize operational v/c ratios of all critical lane groups. Webster developed a formula that introduces the minimum cycle length for isolated intersections by using the strategy [8]. He assumed the equal v/c ratio of 1.0 [8]. Pignataro suggested designing cycle lengths by equalizing v/c ratios by referencing the Webster's equation [9]. Later, McShane and Roess [5] and the HCM [1, 7] suggested determining cycle lengths by balancing v/c ratios of critical lane groups with the user-specified v/c ratio, rather than one. The benefit of this strategy includes the support from the explicit and easy-to-follow manual computation procedure. It does not require a complicated optimization searching method.

2.2.2 Minimize Aggregated Delay to All Vehicles

A strategy of aggregated delay minimization is to design signal-control parameters by minimizing aggregated delay of all vehicles. Many existing computer-based models utilize this group of signal timing design strategies. The computer-based techniques have been utilized for the strategy, since the manual computation is limited to finding the optimal signal timing parameters due to the complexity of minimizing either the aggregated delay or the disutility function containing the aggregated delay, stop and fuel consumption [1, 7].

Various computer-based traffic models were developed with a strategy of minimizing their disutility functions [10, 11, 12]. Such models utilize their own performance-estimation methods. For example, TRANSYT-7F (version 8) uses the 1997 HCM delay equation [12], and SYNCHRO uses a percentile delay model [11]. Those programs compute delay based on the saturation flow rates initially specified by users throughout the design process. Although the initial saturation flow rate can be estimated with the HCM method, the procedure leaves room for error when permitted left turns are

involved. This is because the saturation flow rate of a permitted left turn changes when signal timing changes. Although most of the intersections in the nation allow permitted left turns as default, the models do not explicitly design the signal timing parameters by adjusting saturation flow rates for each signal timing alternative. As a result, there is a question whether their signal timing is truly optimal when the models are evaluated with the HCM procedure.

Saito and Fan [13] attempted to find signal control parameters that minimize aggregated delay using the HCM performance evaluation procedure. They trained an artificial neural network through a supervised learning process to make the network capable of providing the HCM results. They used the neural network as a performance evaluation tool and searched the optimal signal control parameters based on a tree-searching method. Since the main objective of their study was to compare different searching methods, no evaluation of the HCM optimal solution was done. It should be noted that the HCM delay was roughly estimated in their work based on their trained neural network by accepting anticipated errors.

The aggregated delay minimization strategy may yield unbalanced LOS at intersections. The critical-lane groups on a minor street would experience higher delays than the ones on a major street. This is because traffic demand is usually biased over the streets. When such delay differences become severe, unbalanced LOS will result.

2.2.3 Equalize Control Delays of Critical Lane Groups

The HCM suggests balancing the LOS of critical lane groups when traffic volume is highly biased over lane groups. To balance the LOS, control delay of all critical lane groups should be equalized since the HCM defines control delay as the measure of effectiveness for isolated intersection facilities [1, 7]. Although balancing the LOS

would increase aggregated delay, it would promote a balance in the quality of service among all critical lane groups at isolated intersections.

The HCM suggests conducting a set of LOS studies interactively, but no specific guideline is provided [1, 7]. No literature or current computer-based models using this strategy are available.

2.3 Average Green Time Estimation for Traffic-Actuated Phases

For the traffic-actuated system, engineers have directed their primary concerns toward better estimations of average green time and average cycle length to evaluate the system's performance. Earlier editions of the HCM suggested evaluating the performance of a traffic-actuated system based on an assumption that traffic uses its green time under the traffic-actuated operation as efficiently as 95 percent of saturation [14, 15]. The HCM proposed using 0.95 as an actual v/c ratio for actuated systems. In 1995, however, Akçelik found that the actual v/c ratio of traffic-actuated system is less efficient than that [16, 17].

Akçelik and Chung developed a regression model based on simulation data to estimate the operational v/c ratio of actuated phases [17]. They estimated the relationship between volume-to-saturation (v/s) ratio and v/c ratio. It was shown that the saturation of an actuated green is as low as 0.4 when approach volume is low. In the same year, Akçelik [16] developed an analytical model that estimates the average green times and cycle lengths of traffic-actuated controllers by utilizing the regression model developed by Akçelik and Chung. He designed an iterative computational structure to estimate the average green time to explain the relationship between the green and red phases [16]. First, he estimated (1) the operational v/c ratio and (2) the expected number of vehicles served during a phase based on the operational v/c ratio estimated. Then, with the

expected number of vehicles, he computed average green time and operational v/c ratio for the following stage. He suggested following those steps to compute the average green time in the final stage.

In 1996, Courage et al. developed a new approach to analyze the capacity of a traffic-actuated system by dividing the green time of an actuated phase into two different portions: queue service time and phase extension time [18]. This work was undertaken in connection with the National Cooperative Highway Research Program (NCHRP), Project 3-48, and will be referred to in this dissertation as the NCHRP method.

Queue service time is the time required to serve the queue accumulated at the stop line. Courage et al. proposed estimating the queue service time with an assumption of fully saturated operation (v/c = 1.0) based on queue accumulation polygons. Phase extension time is the duration that begins from the end of queue service time and ends at the green termination caused by either failure of successive vehicle detection or the full use of maximum green time. They proposed to estimate the phase extension time with an analytical model based on the bunched exponential distribution function. The computational procedure they proposed uses an iterative computation due to the interrelation among green, red and cycle length. They compared the results from their model to the simulation and field data and found that the R-square values to those are 0.90 and 0.95, respectively. Due to their models, the capacity-analysis procedure for the actuated system becomes as surefooted as the one for pretimed control. Potential improvements to the NCHRP method will be explored in Chapter 4 of this dissertation.

For the capacity and level-of-service analysis of traffic-actuated systems, the versions of the HCM published in the years 1997 [1] and 2000 [7] suggested using the

procedure developed by Courage et al. Although the HCM traditionally has utilized straightforward worksheet procedures, it employs the iterative computational structure by recognizing the complexity of traffic-actuated operation and computer technologies presently being used.

2.4 Maximum Green Time Optimization

Several researchers attempted to construct guidelines for designing optimal maximum green time parameters for traffic-actuated controllers. Kell and Fullerton [19] showed that traffic engineers had arbitrarily determined the maximum green time in practice within a range of 30 to 60 seconds. They developed a rough relationship between maximum green parameters and optimal green time designed for pretimed operation. They recommended that the maximum green parameters of traffic-actuated systems should be 1.25 to 1.50 times longer than the pretimed optimal green times. However, their method provides ambiguity since the value of the multiplying factor is not explicitly defined but ranges from 1.25 to 1.50.

In 1985, Lin [20] verified the relationship between the average delay and the maximum green parameters based on simulation data. He conducted a set of exhaustive simulation analyses through RAPID, a computer-based simulation program developed for the study, by considering various traffic and control conditions. Then, he developed a linear relationship between the optimal maximum green parameters and the optimal pretimed green. He suggested guidelines that could be used to design maximum green parameters based on different levels of the peak hour factor (PHF) (see Table 2-1).

Lin suggested designing the maximum green parameters by computing the optimal pretimed green times. When PHF is close to one, he suggested setting maximum

Table 2-1: The maximum green time design guideline suggested by Lin.

PHF	1.00	0.85	0.70
Maximum green time	$G_{PRE}^{(1)}$ + 10 sec	G_{PRE} × 1.8	G_{PRE} × 2.5

[1] G_{PRE} : optimal green time for pretimed operation

Note: Add 10 more seconds for the right turns.

green by adding 10 more seconds to the green time designed for the pretimed operation. When PHF is close to 0.85 or 0.7, he proposed multiplying the green time designed for pretimed operation by 1.8 or 2.5, respectively. He also tested the elasticity of the maximum green by varying up to 20 seconds from the optimal value. The results showed that average delay rapidly increases when a maximum-green parameter becomes shorter than the one designed for the pretimed operation. When the parameter becomes longer than the one designed for the pretimed operation, the average delay increases, but the amount of increase is insignificant. He recommended using longer maximum green time when uncertainty exists. The method he proposed is relatively easy to implement. However, it still provides ambiguity, in model selection, because of the range of the PHF.

In 1989, Courage, Luh and Wallace [21] investigated the optimization of traffic-actuated signal timings by utilizing computer-based models called the Signal Operation Analysis Package (SOAP) and Network Simulation (NETSIM). They reflected a higher operational v/c ratio (0.95) and a shorter traffic-analysis period (15 min) in their study and concluded that the maximum green times should be somewhat shorter than proposed by Kell, Fullerton and Lin. Their research showed a delay-changing trend similar to the one found by Lin; the average delay rapidly increases when a maximum green shorter than the optimal is used, but the delay slowly increases when the maximum green becomes longer. It also was indicated in their conclusion that the maximum green

parameters have little or no impact on the performance of the traffic-actuated system when traffic volume is low, while they become significant when traffic volume increases.

In 1993, Orcutt [22] suggested setting maximum green time long enough to serve 1.3 times the average queue length in order to accommodate the peak volume during the period of maximum green. He proposed a "realistic but fictitious" cycle length for the maximum green computation and then divided the cycle length for phases based on the pretimed signal timing design method. However, he failed to determine the most efficient cycle length for the traffic-actuated system. It also was proposed that maximum green should be in the range of 20 to 30 seconds and should not be shorter than 15 seconds. He suggested that there be at least a 5-second gap between minimum and maximum green times.

2.5 Computer-Based Traffic Analysis and Design Models

Computer-based models are developed to assist traffic engineers in the performance evaluation of isolated intersections and the design of signal timing parameters. Such models include Enhanced Value Iteration Process Actuated Signals (EVIPAS), Signalized and unsignalized Intersection Design and Research Aid (SIDRA), Corridor Simulation (CORSIM), SOAP and SYNCHRO.[1] Those models were reviewed and are summarized below.

In 1987, Bullen, Hummon, Bryer and Nekmat [23] introduced EVIPAS, a simulation program that estimates the performance of traffic under a traffic-actuated system by utilizing a linear car-following model. The program was developed based on its parent model, Value Iteration Process Actuated Signal (VIPAS). The EVIPAS model

---

[1] The name SYNCHRO stands for "signal and synchronization."

is capable of optimizing a wide range of traffic-actuated control parameters including maximum green, minimum green, unit extension, added-time-per-vehicle actuation, maximum initial time, minimum gap, time before reduction, and time to reduce to minimum gap for any or all phases of a dual-ring controller [24]. When designing actuated control parameters, it simulates the entire set of systems defined based on all feasible combinations of control parameters, and then it selects the best. It is the best existing model that explicitly designs the signal control parameters for traffic-actuated operation. Although the model has been tested and validated, it has not been widely used because it requires (1) too-detailed input data and (2) a very long execution time (exhaustive searching) [18]. Such inefficiencies of the program have prevented traffic engineers from employing the model to design signal timing.

The SYNCHRO program was developed by Trafficware in California and is capable of optimizing signal timing parameters for both pretimed and traffic-actuated systems by utilizing the percentile-delay model. Percentile delay is the average delay of a set of expected estimated delays, based on various vehicle-arrival conditions. The delay model is based on an assumption that the vehicle-arrival pattern follows the Poisson distribution function [25, 26]. Signal timings are designed with SYNCHRO by minimizing the disutility function that includes percentile delay, stops, queue length and vehicles unserved as the disutility factors [25, 26, 27, 28]. It employs models that estimate the effects of phase skipping and time-to-gap-out for the performance estimation of a traffic-actuated system.

Cycle length and splits are designed in two different steps with SYNCHRO. When SYNCHRO optimizes cycle length, constrained by the minimum and maximum

lengths, two possible design strategies can be taken. One is to provide optimal cycle length that can serve the critical percentile vehicle arrivals, and the other is to minimize its disutility function [28]. The former is used when traffic is undersaturated. Ninetieth percentile arrival is used as the critical volume when cycle length is less than 61 seconds. Seventieth percentile arrival is used as the critical volume when cycle length is between 60 and 91 seconds. When cycle length is greater than 90 seconds, fiftieth percentile vehicle arrival is used [11]. The latter is used when the v/c ratio reaches close to or over one.

The program optimizes the split based on the critical number of vehicle arrivals. To find the critical volume, the program checks if the time required to serve the ninetieth, seventieth and fiftieth percentile vehicle arrivals (during a cycle) can be accommodated into the cycle length given. For example, when the given cycle length is not able to handle the ninetieth percentile volume, the program begins to check with the seventieth percentile vehicle arrivals and then with the fiftieth percentile. If extra time is left after this process, SYNCHRO assigns it to the major street phase. It should be noted that SYNCHRO does not explicitly consider queue-service time as a minimal phase length but considers the time required to serve the target percentile volume. Therefore, a longer minimal phase length is defined by SYNCHRO compared to the one from the HCM. For traffic-actuated controllers, SYNCHRO determines the maximum-green parameters by splitting a cycle length.

The SIDRA model was developed in Australia and has a functionality to analyze the performance of a traffic-actuated system by estimating the average green time and cycle length based on a given set of control parameters [10]. This model estimates the

average green time of an actuated phase by utilizing a regression model to estimate the expected operational v/c ratio based on the maximum green and unit extension parameters. Maximum green parameters are required as input. When maximum green parameters are not specified as input, SIDRA assigns 50 and 20 seconds by default to the parameters for major and minor movements, respectively. The SIDRA model does not optimize the maximum green.

CORSIM is a simulation program developed through various research studies under the supervision of the Federal Highway Administration (FHWA) over a period of twenty years. It is a microscopic model that simulates traffic on freeway and network facilities controlled by various types of traffic controls. The program is also capable of simulating traffic-actuated systems [6, 29]. It has been validated and widely implemented in various traffic studies. The model does not provide any optimization routines for signal timing parameters, but it provides reliable simulation results.

The Signal Operations Analysis Package (SOAP) is a model that designs and evaluates control parameters for isolated intersections operated with either pretimed or actuated controllers [30]. It was introduced in 1984 and has been widely used in many traffic studies [18]. The model designs the signal control parameters by minimizing its disutility function, a linear sum of each approach delays and stops. When SOAP designs signal timings for traffic-actuated systems, it uses one second of analysis step size and 0.95 target v/c ratio as default. For pretimed control, five seconds of step size is used. The model does not optimize maximum green parameters explicitly but estimates them through the pretimed signal timing optimization process.

A brief history of the evolution of SOAP is relevant to this dissertation because the algorithms to be developed in Chapters 3 and 4 will be embodied in an enhanced version of SOAP. The name has been used since 1976 to represent the most current version of the signalized intersection timing software offered by the Transportation Research Center at the University of Florida. The chronology is as follows:

- 1976: The first version of SOAP was produced as a mainframe software product that accommodated traffic volumes from 48 consecutive time periods. Its main focus was on pretimed electromechanical controllers with multiple "dial" operation. Separate routines were developed to implement the computations for a single approach and a single time period on a programmable hand-held calculator of 1976 vintage.

- 1980: SOAP/M was developed as a single period analysis program compatible with the APPLE II microcomputer.

- 1984: SOAP84 was developed using the original FORTRAN code to create a PC (MS-DOS)-compatible version of this software product.

- 1990: A single period version, simply called SOAP, was developed from the original code and implemented as a component of the WHICH traffic model integrator for isolated intersection models.

- 2001: SOAP2K will be developed as a part of the study described in this dissertation. The development and testing activities are described in Chapter 5.

2.6 Search Techniques

Several searching algorithms were reviewed for the development of a computational procedure that is capable of finding the best signal control parameters

among the feasible combinations under a given signal timing design strategy. Genetic algorithm and hill-climbing search methods are described in the following subsections.

2.6.1 Genetic Algorithm (GA)

GA is a heuristic searching algorithm that mimics a human being's evolution process, and it is well known as one of the representative machine learning process [31]. The computational structure of GA is illustrated in Figure 2-3. It deals with a set of populations, each one of whose members has its own chromosome by representing a single iteration as a generation. In each generation, offspring and mutants are produced from the parents in the population, and the members who fail to adjust themselves into the environment are weeded out, based on their fitness, as measured by an objective function, to the environment. After a number of generations, the population contains only the best members who are optimally suited to the environment. They are the ones that optimize the objective function.
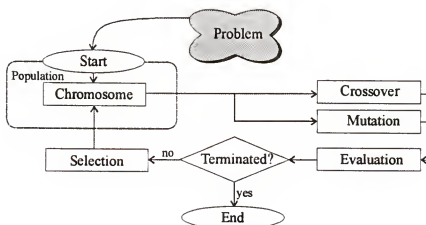
Figure 2-3: Genetic algorithm

The algorithm performs multiple directional searches by maintaining a population of potential solutions whose chromosomes are represented as sets of binary numbers. The crossover and mutation modules in the algorithm introduce offspring and mutants into the system by replacing the parents. Their working procedures are illustrated in Figure 2-4 (a) and (b). In these figures, chromosomes are represented by binary strings. The crossover and mutation points are determined, based on random numbers drawn.

(a) Crossover process of GA

(b) Mutation process of GA

Ci : Chromosome i
Pi : Evaluated performance (objective function value) of ith chromosome

(c) Selction process of GA
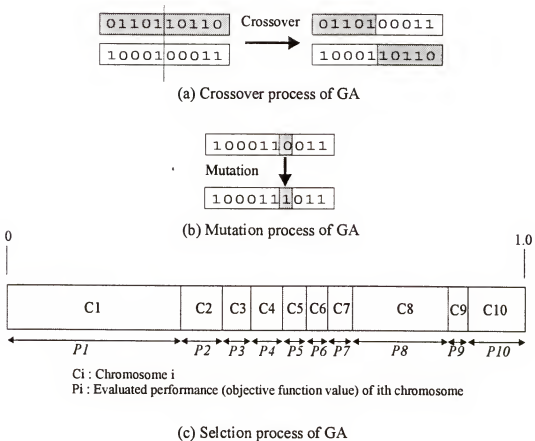
Figure 2-4: Genetic algorithm modules

At each generation, the performance of each binary string, an individual in the population, is evaluated. Also, it is determined whether the population has reached its optimal. If so, no further computation is performed. If not, the population is adjusted by the selection module based on the evaluation function values of the chromosomes.

Figure 2-4 (c) illustrates the general selection module designed for the maximization problem with 10 population sizes. Selection is performed by drawing ten random numbers which range from zero to one, based on the ratio (Pi) of the objective function value of the ith chromosome to the sum of objective functions of all chromosomes in the population. This yields a chromosome providing a higher objective function value, and therefore a higher probability to be selected for the next generation in the maximization problem.

During the past few years, GA has been studied and implemented in a number of engineering fields and has gained much attention from the optimization community [31, 32, 33, 34]. Traffic engineers also have implemented GA to optimize traffic control parameters. Abu-Lebdeh and Benekohal [35] optimized pretimed signal control parameters for oversaturated flow conditions by applying GA.

Hadi and Wallace [36] implemented GA to optimize a set of traffic-control parameters and phase sequences for pretimed control systems. Park, Messer and Urbanik [37] utilized GA to optimize offsets of two consecutive oversaturated intersections.

GA also is capable of detecting multiple solutions if they exist. Park, Messer and Urbanik [37] showed that there would be a set of signal timings that provides identical objective-function values (multiple solutions). Providing such options to traffic

engineers would be beneficial because they can select the best among the possible timing and phase sequences based on the conditions of adjacent intersections.

2.6.2 Hill-Climbing Search

Hill-climbing and best-first searches are categorized as blind search methods since they search for solutions without knowing the direction to move in the next step [38]. In maximization problems, for example, those methods move to the optimal solution by always facing toward to the highest point. When no movement can be made further, those methods return the point as the optimal solution.

In a hill-climbing search, the searching direction to the solution is always determined by continuing to select the best child from the best parents listed. The algorithm of hill climbing is summarized below:

Step 1: Set a list of the initial nodes in the problem and sort them by their objective function values. Nodes expected to be close to the goal should be prior to others that are farther from it.

Step 2: Let n be the first node on L.

Step 3: If n is the goal node, stop the computational procedure. If not, replace n with its best child among the children.

Step 4: Remove n from L. Sort the children of the point n by their objective function values and replace the best one with the point n. Add the new n to the front of L and go to L.

The hill-climbing method searches for a solution with depth-first favor, while best-first search does so with breadth-and-depth first favor [38]. The best-first search method is basically similar to the hill-climbing method except for the rule of moving-

direction selection. The best-first search method moves toward the direction determined based on both child nodes and the other nodes listed in L.

These searching methods guarantee to find the optimal point in a feasible region. However, they have a problem with local optimal points since they provide no capability to escape from them to the global optimal [38]. Various techniques such as simulated annealing have been introduced to improve this deficiency by increasing the convergence time of those methods [31]. However, the limitation has not been removed entirely although the possibility of falling into the local optimum was reduced.

CHAPTER 3
DEVELOPMENT OF THE COMPUTATIONAL PROCEDURE FOR THE HCM-BASED SIGNAL TIMING DESIGN

This chapter describes the development and testing of a computational procedure that seeks the best set of signal timings based on the HCM performance evaluation methodology. It was pointed out in Chapter 1 that no such procedure currently is documented in the literature. This discussion includes a description of the structures and models used in the proposed computational procedure. Specifically, the following topics are addressed:

1. The functional requirements for the proposed procedure were established at the initial stage of the research. The procedure must accommodate two signal timing design strategies: (a) minimization of the aggregated delay to all vehicles and (b) equalization of delay among all critical lane groups. A third commonly applied strategy, generally referred to as v/c ratio equalization, was excluded from this study because of its well-documented analytical models [1, 5, 7, 19].

2. The structure of the computational procedure was designed to meet the functional requirements. Two optimization formulations were developed for the selected strategies. The computational procedure developed in this study was implemented to find the best signal timings under those formulations.

3.   The specific computational searching process was developed within the context of this structure using advanced search techniques. The computational details were configured to increase the efficiency of searching by avoiding an exhaustive search in a whole set of feasible signal timings.

4.   Two simple test runs were made with hypothetical intersections to verify that the computational procedure was able to find the "best" HCM-based solution and that the signal timings obtained by the proposed method are different from those designed by the existing computer-based models.

3.1 Requirements of the Computational Procedure

To guide the development of the computational procedure, two sets of requirements have been specified: functional and computational requirements. They are described in the following subsections.

3.1.1 Functional Requirements

Functional requirements are the ones that the computational procedure should be capable of achieving. They include

1.   optimization of pretimed cycle length and splits that minimize aggregated delay of all lane groups;

2.   determination of pretimed cycle length and splits that equalize control delays among all critical lane groups;

3.   optimization of the maximum green time parameters that minimize the aggregated delay;

4.   improvement in the NCHRP method that estimates the average phase length of a traffic-actuated phase based on given traffic volumes and traffic-actuated control parameters;

5.   recognition of vehicle and pedestrian minimum green times, explicitly;

6.   compatibility with the dual-ring concurrent phasing scheme; and

7.   limitation of input data (the computational procedure should not require additional input data other than those required by the HCM, except minimum and maximum green time, which are traffic-actuated control parameters excluded from the current HCM input data set).

The strategy of control-delay equalization for critical lane groups is used instead of the strategy of LOS equalization for critical lane groups. Although the HCM uses the terminology of the LOS equalization strategy, the use of the LOS threshold values in design procedure would be insignificant. The measure of effectiveness of the LOS is of more significance than the LOS values themselves. It is required that the computational procedure be based on the dual-ring concurrent phase scheme since the scheme is able to represent any phase sequences applicable in both pretimed and traffic-actuated controllers.

3.1.2 Computational Requirements

The development of the computational procedure should be based on the computational requirements listed below. They include

1.   an efficient optimization algorithm that avoids an exhaustive search, and

2.   incorporation into a software environment that promotes productive application; computer-based development should be done for efficient tests of the computational procedure.

3.2 The Structure of the Computational Procedure

The structure on which the computational procedure is developed consists of two major parts: evaluation and searching. Since the best set of signal timings is to be found based on the HCM evaluation method, the HCM evaluation should be made to assess each alternative set of signal timings during the searching procedure. In addition, the proposed computational procedure should be incorporated into a software environment for the efficiency of the tests as specified in functional requirements. This implies the development of the HCM computational engine being capable of handling all possible combinations of geometric and control conditions as specified in the HCM. As an evaluation tool of the computational procedure, the Highway Capacity Software (HCS) is employed and incorporated with the proposed computational procedure. The HCS program was developed at the University of Florida with the support of the FHWA and TRB for the automation of the HCM evaluation procedure. It has been widely implemented in the nation for the automation of the HCM evaluation procedure [2, 18]. Employing HCS avoids consuming the tremendous amount of time required to develop an evaluation module, while one is already available at hand. Since HCS is the computer version of HCM, it covers all combinations of various geometric and control conditions interested in the computation.

The searching technique that promotes the best signal timings and incorporates into the computational structure using HCS as an external module has been developed. Various searching techniques have been nominated and evaluated. Such techniques include both heuristic and algorithmic searches. Brief descriptions on those methods are included in this section, based on the compatibility in the computational structure and the features of the computational procedure.

The requirements of the computational procedure include the use of the dual-ring signal timing scheme (see Figure 1-2). The six decision variables were designed to represent the eight different phase lengths in the dual-ring phase scheme. Six decision variables, instead of eight, were declared in order to reduce the computation time (long computation time is generally required with the large number of decision variables). The configuration of the six decision variables is presented in Figure 3-1. The first variable, x_1, specifies the cycle length, and the second variable, x_2, represents the length of a phase pair in the left side of the barrier. The other variables, x_3, x_4, x_5 and x_6, represent the lengths of the phases 2, 4, 6 and 8. The phase length includes green and intergreen time, consisting of yellow and all red times. Based on this variable structure, the nominated searching techniques were evaluated.
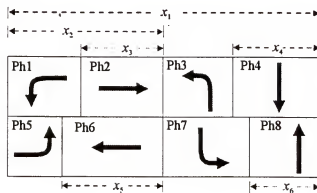
Figure 3-1: Independent variables designed for the dual-ring phase scheme

The blind search techniques including breadth-first, depth-first and A* searches were nominated and rejected. The blind search methods require a tremendous size of searching tree for the dual-ring phase scheme since one node at the tree should generate twice the number of child nodes than the number of decision variables at each depth [38].

For example, with the six decision variables designed for the dual-ring phase scheme, the blind searching method should be based on a tree (searching tree) consisting of 12^n nodes at depth n. One depth represents a unit alternation of a single decision variable. In addition, the size, n, per se cannot be determined because of the undeterminable depth.

Frank-Wolf algorithm [39] was nominated and rejected. It turned out that the computational feature of the non-linear operations research algorithm did not fit into the proposed structure that uses HCS as an external evaluation module. This conventional operations research method finds the optimal solution by the guidance of a set of the first-order partial-differential equations of the objective function. Due to the feature of the algorithm, it does not require the use of HCS as an evaluation tool. It requires the development of a new external module that generates a set of partial differential equations, and the module should cover all geometric and control conditions specified in the HCM. In addition, it is not an easy task in this case to obtain the first-order partial-differential equations of the objective function, which would contain a set of the HCM delay equations. Although a set of partial differential equations can be obtained, this conventional operations research method (1) makes the whole computational procedure much more complex and (2) decreases the efficiency in terms of development due to the external module that should be built to assist the searching procedure.

GA was nominated and considered as the best searching model that fits into the proposed computational structure. GA is a heuristic searching method that mimics the evolution process. Figure 3-2 presents the proposed computational structure including the GA and HCS portions of computation. HCS was used in the system performance evaluation part, and GA was assigned to check the optimality and to update the signal-

timing parameters. GA provides the flexibility in its objective function and constraints [33]. It mostly finds the global optimal solution with any shape of the objective function, while the conventional non-linear programming method requires the changes of a whole computational algorithm. For example, it does not require any change in the computational procedure with the different signal timing design strategies but just requires the replacement of the objective function.
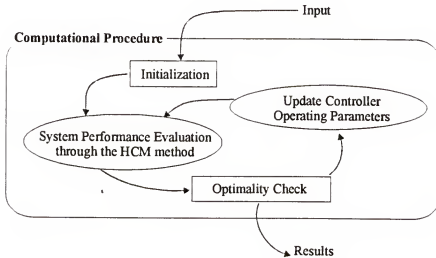
Figure 3-2: Proposed structure of the computational procedure

The hybrid GA (HGA) is finally selected and implemented into the proposed computational procedure. Although the conventional GA mostly guarantees to find the global optimal solution with the large number of generations (iterations), in worst cases such number of generations would result in a long convergence time. With a small or moderate number of generations, the solutions provided by GA would not be global optimal. In addition, with a small or moderate number of generations, the final solutions

would be different when different initial random seed numbers are used. In order to overcome those two shortcomings of the conventional GA, HGA was implemented.

HGA is the combination of GA and the scaled hill-climbing search method. The hill-climbing search is a heuristic method that finds the optimal point by moving, with a given step size, from any point in the feasible area to the direction that introduces better value of the objective function. When no point provides a better value of the objective function than the point with a flag, the hill-climbing search returns the flagged point as the optimal solution. However, this method has a limitation on escaping from the local optimum [38]. By using HGA, it is expected that the GA should return a stable output solution with a reduced convergence time due to the assistance of the hill-climbing method. Also, it is expected that GA should eliminate the local optimum problem of the hill-climbing method. The structure of HGA is presented in Figure 3-3.
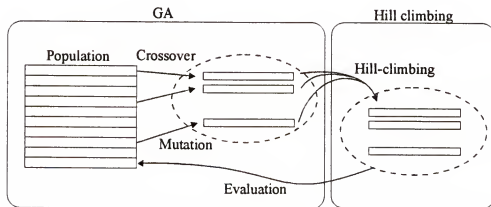
Figure 3-3: The structure of HGA

3.3 Formulation Development

A set of optimization problems was formulated for the signal timing design strategies selected for the study: the aggregated delay minimization and the control delay equalization. Two minimization problems are presented in this section. The formulations include two different objective functions and a single set of constraints, which were applied to both minimization problems.

3.3.1 Minimization of Aggregated Delay of All Vehicles

An objective function that represents the average control delay of all vehicles entering an isolated intersection is developed and proposed for the strategy of aggregated delay minimization. It is expected to minimize the aggregated delay of all vehicles by minimizing the average delay of those vehicles. The optimization formulation set for the strategy is presented below.

$$\text{Minimize} \qquad \frac{\sum_{i=1}^{8} v_i d_i'(x_1, \cdots, x_6)}{\sum_{i=1}^{8} v_i} \tag{3-1}$$

$$\text{subject to} \qquad x_1 \geq x_2, \tag{3-2}$$

$$x_2 \geq Max\left[\sum_{i=1}^{2}\left(G_i^{\min} + I_i\right), \sum_{i=5}^{6}\left(G_i^{\min} + I_i\right)\right], \tag{3-3}$$

$$x_1 - x_2 \geq Max\left[\sum_{i=3}^{4}\left(G_i^{\min} + I_i\right), \sum_{i=7}^{8}\left(G_i^{\min} + I_i\right)\right], \tag{3-4}$$

$$x_2 - x_i \geq 0 \qquad for\ i = 3\ and\ 5, \tag{3-5}$$

$$x_1 - x_2 - x_i \geq 0 \qquad for\ i = 4\ and\ 6, \tag{3-6}$$

$$x_2 - x_{3+(i-1)/2} > G_i^{\min} + I_i \qquad for\ i = 1\ and\ 5, \tag{3-7}$$

$$x_1 - x_2 - x_{3+(i-1)/2} > G_i^{\min} + I_i \qquad for\ i = 3\ and\ 7, \tag{3-8}$$

$$x_i > G_i^{min} + I_i \qquad \text{for } i = 3, 4, 5 \text{ and } 6, \qquad (3\text{-}9)$$

$$x_i - C_{min} \geq 0, \qquad (3\text{-}10)$$

$$C_{max} - x_1 \geq 0, \text{ and} \qquad (3\text{-}11)$$

$$x_i \geq 0 \qquad \text{for } i = 1, \cdots, 5 \text{ and } 6. \qquad (3\text{-}12)$$

where

x_i	=	decision variables (see Figure 3-2),
v_i	=	traffic volume of the movement assigned to the dual-ring phase i,
d_i	=	control delay of the movement assigned to the dual-ring phase i,
G_i^{min}	=	minimum green time of the dual-ring phase i.,
I_i	=	intergreen time of the dual-ring phase i.,
C_{min}	=	minimum cycle length and
C_{max}	=	maximum cycle length.

A set of constraints is designed to guide the decision variables to be in the feasible area, which is equivalent to the dual-ring representation scheme. Equations 3-2, 3-3, 3-4, 3-10 and 3-11 control the combination of phase lengths for each side of a barrier and a cycle length. Equations 3-5, 3-6, 3-7, 3-8, 3-9 and 3-12 control individual phase length.

3.3.2 Equalizing Control Delays of All Critical Lane Groups

Equalizing control delay strategy provides a balanced LOS for all critical lane groups. As indicated, the measures of effectiveness (control delays) of the HCM LOS are used in the proposed computational procedure as control parameters instead of the LOS that the HCM defines. This is because LOS is a qualitative category that describes the operational condition of facilities but not an index indicating a meaningful operational condition.

A minimization problem was formulated for the strategy of control delay equalization (see Equation 3-13). The objective function consists of two major portions. The first one, the first term in the function, minimizes the standard deviation of control delays of critical lane groups. By minimizing the standard deviation, it is expected that the control delays of critical lane groups should be equalized. The second portion, the second and third terms, was designed to find the minimum among the multiple solutions with the zero standard deviation. A scale factor, α, was provided to scale the effect of those two major portions. The objective function developed for the strategy is presented below.

$$\text{Minimize} \qquad s + |1 - \alpha| \cdot \frac{\sum_{i=1}^{8} v_i d_i / \sum_{i=1}^{8} v_i}{\sum_{i=1}^{8} v_i d_i / \sum_{i=1}^{8} v_i + 1} + \alpha \qquad (3\text{-}13)$$

where

C_{max} = maximum cycle length,

s = standard deviation of control delays of all critical lane groups,

v_i = traffic volume of the movement assigned to the phase i and

α = a control factor.

The control factor, α, is obtained by the following equation:

$$\alpha = \begin{cases} 0 & \text{if } s \leq 1 \\ 1 & \text{if } s > 1 \end{cases} \qquad (3\text{-}14)$$

The standard deviation can be decreased to zero when it is minimized. When the standard deviation is larger than one, the second term becomes zero and the third term

remains one. In such a case, control delay equalization is only active. When the standard deviation becomes smaller than or equal to one, the second term, ranging between zero and one, becomes effective while the third term becomes zero. A control factor, α, introduced in the second and third terms controls such conversion of optimization schemes: equalization and minimization.

The constraints presented from Equations 3-2 to 3-12 are applied to both minimization problems since the identical feasible area should be applied. The decision variables in both problems should be in the dual-ring presentation scheme.

3.4 Computational Procedure Development

The proposed computational models used in the proposed HGA searching procedure are presented in this section. The GA and hill-climbing search methods require the mathematical definition of their computation modules. The following sections describe the modules developed for those two searching methods.

3.4.1 Modules of GA

The conventional GA requires three computational modules: representation, crossover and mutation. GA used in the proposed HGA employs a real number encoding scheme for representation, arithmetic operators for convex crossover and a dynamic mutation operator for mutation.

Representation

The HGA used in the proposed computational procedure employs a real number representation scheme. The conventional GA uses a conventional operator that utilizes a set of binary codes to represent the chromosome of a decimal number. However, it was recommended to use a real coding technique to represent a number for constrained optimization problems [32]. This technique increases the speed of the searching process

by avoiding a converging process required for encoding and decoding binary codes to decimal numbers and vice versa.

Crossover

For the crossover operation, an arithmetical operator was used. An arithmetical operator is the one formulated based on convex set theory [32, 40]. With this operator, offsprings are defined as the weighted average of two parent vectors. Let's denote the parents as v_1 and v_2. Their offsprings are computed as follows:

$$v_1^{new} = \lambda v_1 + (1 - \lambda)v_2, \text{ and} \tag{3-15}$$

$$v_2^{new} = \lambda v_2 + (1 - \lambda)v_1 \tag{3-16}$$

λ is a weight factor. By setting the range of λ as $0 < \lambda < 1$, the crossover operation was set to be in a convex hull. When the range of λ is defined differently from that, affine and linear crossover can be implemented. Affine crossover is the one introducing a child from the outside but the extended line of the convex hull. Linear crossover is the one introducing a child from the area that can be represented by a circle. The convex crossover was employed in this study by setting λ as 0.38, which introduces the "golden section" (1:1.618) of the convex hull.

Mutation

Dynamic mutation was employed. This operator is designed for fine tuning capabilities aimed at achieving high precision. Two operators are defined (see Equations 3-17 and 3-18). When mutation occurs, the active operator is selected from those two based on a random number drawn. Let's denote the selected element for mutation as x_k. The equation used for mutation should be either

$$x_k^{new} = x_k + \Delta\left(t, x_k^U - x_k\right) \text{ or} \qquad (3\text{-}17)$$

$$x_k^{new} = x_k - \Delta\left(t, x_k - x_k^L\right) \qquad (3\text{-}18)$$

where

x_k^U =   upper bound for x_k,

x_k^L =   lower bound for x_k,

$\Delta(t, y) =$ $y \cdot r \cdot (1 - t/T)$,

r =   a random number within the range of $[0,1]$,

t =   generation (iteration) number and

T =   target generation number.

The function Δ returns a value within the range of $[0, y]$, and the value approaches zero when t increases. It makes the mutation operator generate a mutant that can be anywhere in a whole feasible area when the number of iterations is small. When the number of iterations becomes large, mutation occurs locally.

3.4.2 Step Size of the Scaled Hill-Climbing Search

In the hill-climbing search, its step size plays an important role. When the step size is small, the expected number of iterations required to reach the optimum would be increased, and more chances to converge to the local optimum would be provided. On the other hand, when the step size is large, the searching speed becomes fast, but it only provides a coerce estimation of the optimal solution.

For efficient performance of the hill-climbing search, the scaled hill-climbing search method was developed for this study. In the scaled hill-climbing search method, the step size was designed to vary over the searching process. It was intended to search

the feasible space with the large step size at the initial stage, and the size becomes smaller than at the latter stages. This changing trend of the step size is illustrated in Figure 3-4.
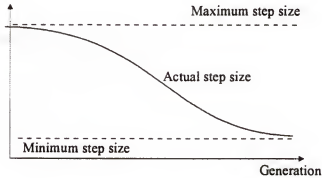
Figure 3-4: Step size used in hill-climbing search

When the searching process begins, a rough estimation of the optimal solution is performed in order to reduce searching time. When the generation becomes large, the step size becomes small so that the searching process will find a finer optimum. An equation is developed to compute the hill-climbing step size dynamically determined during the searching process. Let's denote k be a step size of the hill-climbing search. Then:

$$k = k_{min} + \left(\frac{h - 0.37}{1.63 - 0.37} \right) \cdot (k_{max} - k_{min}) \qquad (3\text{-}19)$$

where

$$h = \begin{cases} 2 - \exp[-(x - u_0)/(u_1 - u_0)] & \text{, when } x < u_0 \\ \exp[-(x - u)/(u_2 - u_0)] & \text{, when } x \geq u_0 \end{cases},$$

k_{min} = minimum step size,

k_{max} = maximum step size,

u_1 = minimum generation number, 1,

u_2 = maximum generation number, target generation and

u_0 = middle point between u_1 and u_2.

The structure of this scaling equation was originally developed for the GA selection procedure as the stabilizing section method [32]. The equation was modified to Equation 3-19 for incorporation into the scaled hill-climbing search method developed in this study.

3.5 Demonstration and Testing

A computational procedure using HCS as an external evaluation module has been developed. Sample tests were done to verify the performance of the proposed procedure. Two hypothetical intersections were designed to test the performance of the procedure for two signal timing design strategies. The proposed computational method was coded with the Visual Basic computer language for the purpose of the tests. A mock HCS program that performs the HCM performance evaluations was developed with the Visual Basic programming language as well as for the evaluation of those two specific test cases. The tests are done to check if the proposed HGA performs reasonably before the complete development of the computational procedure. Based on the results of those tests, the full development of the communication between HCS and the proposed procedure is initiated.

For the GA computation, the population size, the crossover rate, the mutation rate and target generation are set to 10, 0.4, 0.1 and 10, respectively. For the scaled hill-climbing computation, the maximum and minimum step sizes are set to 5 and 1, respectively.

For comparison purposes, two other existing computer-based signal timing design tools were implemented to the test problems: TRANSYT-7F and SOAP. The comparison was made among the signal timings designed by the proposed HGA computational procedure, the HCM based signal-timing design, and those models. TRANSYT-7F and SOAP are the widely accepted computer-based models in practice [18]. The input data files of TRANSYT-7F and SOAP were prepared through the Arterial Analysis Package (AAP) and the Wizard of Helpful Intersection Control Hints (WHICH) [12, 30]. AAP and WHICH are the computer programs that generate different formats of input files for different computer-based intersection-study models. Since the HCM-based signal timing strategies consider the control delay only by excluding the effects from stops and fuel consumption, the stop-penalties of TRANSYT-7F and SOAP were set to zero to make the results comparable.

3.5.1 Minimization of Aggregated Delay

The layout of the hypothetical intersection prepared for the test of the aggregated delay-minimization strategy is illustrated in Figure 3-5. The major and the minor streets have three and one through lanes, respectively, with no turning volume on either approach. The intersection is operated with the pretimed simple two-phase operation. The start-up and ending lost times are set to two and one second, respectively. Ten seconds of minimum green times were used. Other data including saturation flow rate, the percent trucks, area type and PHF are presented in Figure 3-5.

Sets of signal timings were designed based on the proposed method, TRANSYT-7F and SOAP. The signal timings included cycle lengths and the splits (percent of time used for a specific phase). The results are summarized in Table 3-1.
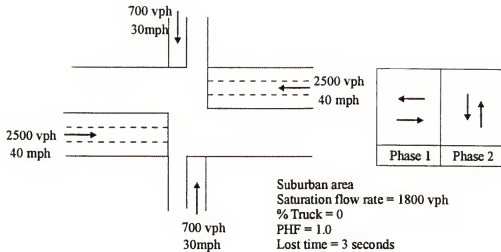
Figure 3-5: The hypothetical intersection used for the strategy of delay minimization

Table 3-2: Optimal cycle lengths and splits designed by different methods

	Optimal Cycle Length	Split	
		Phase 1	Phase 2
TRANSYT	50.0	0.54	0.46
SOAP	55.0	0.56	0.44
Proposed Procedure	94.3	0.57	0.43
Webster's Optimal Cycle	95.0	N/A	N/A

TRANSYT-7F and SOAP suggested 50.0 and 55.0 seconds of cycle length and 0.54 and 0.56 of the split of phase 1, respectively, while the proposed procedure, the proposed method, provides 94.3 seconds of cycle length and 0.57 of the split. The Webster's equation for the optimal cycle length was implemented as well, and 95.0 seconds of cycle length was found as the result. While TRANSYT-7F and SOAP suggested similar values of cycle length and split, the proposed procedure, the HCM-based signal timing method, suggested a similar cycle length, 94.3, to the one from the

Webster's formula, 95.0. The computation time of the proposed method was 8.35 seconds with the Pentium 233 MHz computer.

The comparison of those models with the results only from a single run was not enough to find the differences in those methods. Sets of runs were made to visualize different trends of the intersection evaluation methods embedded in those models. Two different sets of multiple runs were made. First, the cycle lengths presented in Table 3-1 were changed with the fixed the split values, and evaluations were made. The goal was to illustrate the changing tends of the objective function values over different cycles. Second, the splits were varied and evaluated with the subject models with their fixed cycle length presented in Table 3-1. Another set of aggregated delays was obtained and plotted to visualize the changing trend of the objective function value over different splits. The comparisons of the models through those two graphs are presented in Figure 3-6 and Figure 3-7.
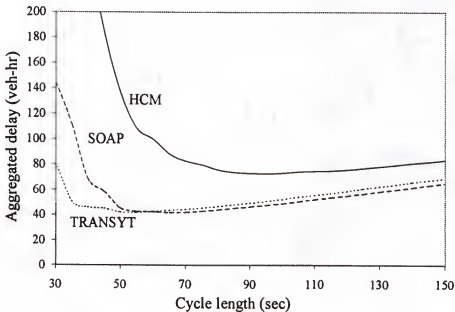
Figure 3-6: Aggregated delay vs. cycle lengths
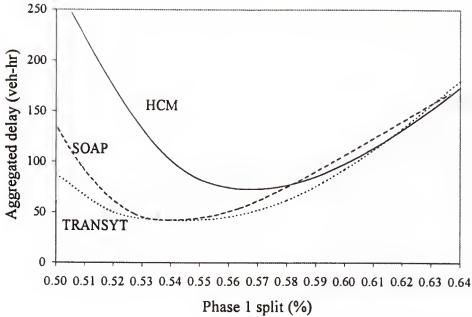
Figure 3-7: Aggregated delay vs. splits

The Figures 3-6 and 3-7 show that the proposed HGA computational method was capable of finding the HCM-based optimal signal timings. The HGA cycle and split values specified in Table 3-1 turned out to be the minimum points of those convex curves drawn based on the HCS runs.

Although the TRANSYT-7F and SOAP programs employ the HCM delay equation in its evaluation function and only the HCM delay was considered in the test by setting the stop penalty parameters to zero, the different changing trends of the objective function values (aggregated delay) were obtained.

The difference comes from the saturation flow rate. The lane-utilization factor was not considered in the estimation of the saturation flow rates provided for TRANSYT-7F and SOAP.

The constant saturation flow rates estimated by the input manager programs, AAP and WHICH, were used with TRANSYT-7F and SOAP. Those input manager programs do not consider the lane-utilization factor in their saturation flow rate adjustment procedure. Such different saturation flow rates lead the results plotted in the Figures 3-6 and 3-7. TRANSYT-7F and SOAP do not consider the saturation flow rate adjusted by all adjustment factors specified in the HCM including the lane-utilization factor.

Figures 3-6 and 3-7 illustrate that the performance evaluation conducted in the TRANSYT-7F and SOAP programs was different from the one suggested by the HCM. This indicated that the signal-timing parameters suggested by TRANSYT-F and SOAP were not the optimal when they were evaluated through the HCM. These figures summarize the current status of traffic engineering practice in signal timing design and the evaluation required by the HCM.

3.5.2 Equalization of Control Delay for All Critical Lane Groups

A hypothetical intersection prepared for the test of the control delay equalization strategy is presented in Figure 3-8. Two crossing streets have a single lane with no turning volume. The intersection is operated with the pretimed simple two-phase operation. The start-up and ending lost times are set to two seconds and one second, respectively. Ten seconds of minimum green time was used for each phase. Other data including saturation flow rate, the percent trucks, area type and PHF are presented in Figure 3-8.

Identical traffic and geometric conditions were provided for all approaches to force the minimal point of the objective function, a convex function, to be at 50% split. This was done to visualize the performance of the proposed procedure with the controlled
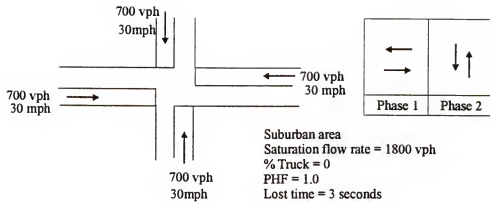
Figure 3-8: The hypothetical intersection used for the strategy of delay equalization

conditions by 50:50 splits and identical traffic volumes. No model currently exists for the delay-equalization strategy. Thus, the proposed HGA procedure was only tested.

The proposed procedure suggested 58.4 seconds of cycle and 50% split after the 8.79 seconds of computation with the Pentium 233 MHz computer. Multiple HCS runs were made by varying the cycle length with fixed 50% splits to visualize the changing trend of the aggregated delay. Although the delay equalization was the main objective, it was also targeted to minimize the aggregated delay among the multiple signal timings that equalize the control delay among the critical lane groups. The results from the multiple runs are plotted in Figure 3-9.

The test results show that the proposed HGA procedure is capable of finding the optimal split, 50% as controlled for the test, and the cycle length. Figure 3-9 shows that the aggregated delay becomes minimal with the 58.4 seconds of cycle length. No
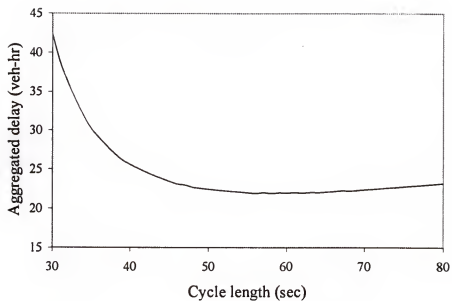
Figure 3-9: Aggregated delay vs. cycle lengths

comparison tests to the other models were done, since no model supporting the strategy of either control delay or LOS equalization exists.

CHAPTER 4
ACTUATED GREEN TIME ESTIMATION AND THE EFFECT OF
THE MAXIMUM GREEN PARAMETER

The average phase length estimation model plays an important role in the maximum green time design procedure. Therefore, improvements in the phase length estimation model should produce better timing plans that will reduce delay to traffic.

This chapter presents the results of the studies conducted (1) to modify the NCHRP method (introduced in Chapter 2) for determining the average length of a traffic actuated phase and (2) to determine if the HCM optimal maximum green time parameters for traffic-actuated operation can be found by a combination of the proposed HGA computational procedure and the modified NCHRP method. The HCM suggests using the NCHRP method for the estimation of the average actuated phase lengths [1, 7]. This method must therefore be viewed as the current defacto standard for average actuated phase length estimation. Improvement of this method would therefore represent a useful contribution to the body of knowledge in traffic control modeling.

The proposed modifications to the NCHRP method include (1) explicit treatment of right turns in lane groups that include both through and right-turning vehicles, (2) modification to the treatment of permitted-left-turns and (3) revisions in the concept of the additional queue-service time.

Using the proposed modifications, the effect of the maximum green time settings on the average delay to vehicles entering the intersection was investigated to determine

whether or not the maximum green time settings could be optimized within a practical range of values. This part of the study was intended to verify the existence of the optimal solution and to determine whether the combination of the proposed computational method and the modified NCHRP method is able to find the optimal solution. The overall average delay was employed as the performance index for the study.

4.1 Improvement Made with the NCHRP Method

For the HCM-based evaluation of traffic-actuated operation, the average lengths of the traffic-actuated phases must be provided as inputs. Improvements were introduced in the study to the NCHRP method for better estimation of the average phase length. They include the through volume adjustment for the shared-through-and-right-turn lane group and the modifications made for the estimation of the permitted-left-turn parameters for the original definition of queue service time. The parameter modifications include the updates based on recent studies and the adjustment in the approaches taken for the modeling procedure. Each of those is described in the following subsections.

4.1.1 Right-Turn Treatment

The NCHRP method suggests using the aggregated volume of right-turn and through movements as the volume loaded on the shared right-turn lane [1, 7, 18]. By using the aggregated volume, it was assumed that all right-turn vehicles should flow on green with the through movement and should generate the vehicle actuation during green. This yields the overestimation of the average phase length due to the failure of counting the right-turn volume turning on red. For example, with the right-turn-on-red (RTOR) operation, the RTOR vehicles turning on red do not affect the vehicle actuation during green. They should be excluded from the lane group volume at least for the average phase length estimation. The HCM indicates that the estimation of the RTOR volume is

a function of many complicated factors and that all right-turn volume should be considered without reduction for the performance estimation [1, 7]. However, it is desirable to exclude the RTOR from the lane group volume for the estimation of the average phase length estimation, which should be based only on the vehicles flowing during green.

In the improvement, four different right-turn treatment cases were categorized, and the shared-through and right-turn lane group volume was adjusted based on the right-turn treatment cases. The right-turn treatments include RTOR, No RTOR, overlap left turn (OLT) and free turn. The permitted right-turn movement and its concurrent movements for each of those treatment types are presented in Figure 4-1. 'Go' indicates the permission to proceed, and 'No Go' indicates prohibition. 'No Conflict' indicates 'Go' without any conflict.

Treatment	Code				
RTOR	1	Go	Go	Go	Go
No RTOR	2	Go	No Go	No Go	No Go
Overlap LT	3	Go	No Go	No Go	Go
Free	4	No Conflict	No Conflict	No Conflict	No Conflict
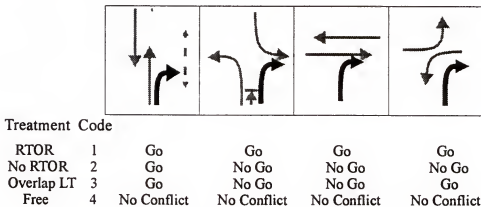
Figure 4-1: Right-turn treatments

In the RTOR case, the RTOR vehicles should be excluded from the lane group volume since they are free from actuations occurring during green. In the No-RTOR case, no right-turn vehicles are permitted to flow on red. Therefore, all right-turn

vehicles should be counted as the lane-group volume without any subtraction. In the OLT case, right-turn vehicles have permission to flow on red concurrently with the left-turn movement turning from the right-hand side approach. The right-turn vehicles flowing on red with the protected left-turn phase should be excluded from the lane-group volume. The OLT right-turn treatment is the popular type of operation where the five-section signal head is used. In the free-turn case, no conflict exists for right-turn vehicles. It includes the exclusive right-turn lane cases. All right turns should not be considered as a through lane group, as defined in the HCM.

For the determination of the right-turn treatment type, a flow chart presented in Figure 4-2 was developed. The procedure was integrated with the NCHRP method. In the NCHRP method, the right-turn treatment type of a shared-through and right-turn lane group volume is determined, and then the lane group volume is adjusted by the following procedure.
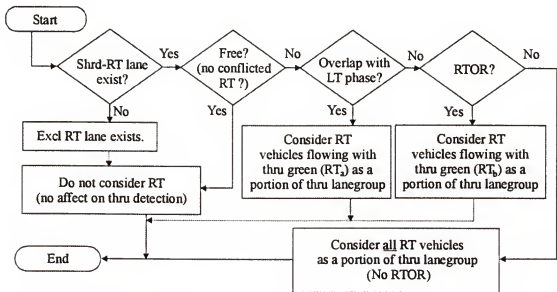
Figure 4-2: Right-turn treatment determination in the modified NCHRP method

For the adjustment of the lane group volume, the estimation of the proportion of right turns turning on red is required for the RTOR and OLT cases. A model that estimates the number of right-turn vehicles turning on red, v_R, was developed. The lane-group volume was then obtained by subtracting the estimated volume, v_R, from the through lane group volume. A set of equations has been developed and used for the estimation of the adjusted lane group volume. The equations are presented below. First, Equation 4-1 presents the one used for the estimation of the adjusted through-lane group volume.

$$v'_{TH} = v_{TH} + v_{RT} - v_R \qquad (4\text{-}1)$$

where

v'_{TH} = adjusted lane group volume,

v_{TH} = the number of through vehicles,

v_{RT} = the number of right-turn vehicles and

v_R = the number of RTOR/ right turns overlapped with left turns.

The value of v_R, the number of right turns that do not affect actuation on green, was stochastically estimated based on the geometric distribution function. The geometric distribution function is the one representing the probability of the number of failures, the number of right-turn vehicles turning on red, expected to be experienced before the occurrence of the first success, the first arrival of a through vehicle blocking a shared lane. The following equations were used for the estimation of v_R:

$$v_R = P_R / (1 - P_R) \qquad (4\text{-}2)$$

$$P_R = v_{RT} / \{P_{RL}(v_{TH} + v_{RT})\} \qquad (4\text{-}3)$$

where

P_R = percent right-turn traffic on the right-most lane and

P_{RL} = percent traffic using right-most lane.

Estimation of v_R requires the estimation of P_R, the percent right turns on a shared lane, which is a function of through vehicles using the shared lane since all right-turn vehicles deserve to flow on the lane. Thus, the estimation of P_R requires the estimation of P_{RL}, the percent lane group using the right-most lane, which is a shared lane. The equations used in the P_{RL} estimation procedure are presented below.

$$P_{RL} = \begin{cases} P_{HL} & \text{, when \# of lanes} = 1 \\ 1 - P_{HL} & \text{, when \# of lanes} = 2 \\ 0.75 - 1.25P_{HL} & \text{, when \# of lanes} = 3 \end{cases} \tag{4-4}$$

where

P_{HL} = percent traffic using most heavily traveled lane.

The HCM lane-utilization factors were utilized for the P_{HL} estimation. The P_{HL} used in this study and the P_{RL} values obtained from Equation 4-4 are summarized in Table 4-1. With either a single- or double-lane case, the value of P_{RL} was directly estimated based on P_{HL}. For the three-lane case, it was assumed that the percent traffic using the second heavily traveled lane should be in the middle of the upper and the lower limits. First, it was assumed that the percent traffic using the second heavily traveled lane could not be higher than the percent of the most heavily traveled lane. Second, it was assumed that it could not be lower than the upper limit of the lowest traveled lane. The upper limit of the percent traffic using the lowest traveled lane would be 50 percent

of the sum of the percent traffic using the second most highly and the lowest traveled lanes. The middle value between its upper and lower limits was used for the estimation of P_{RL}.

Table 4-3: The lane distribution for the right-most lane

Number of lanes	1	2	3
P_{HL} [1]	1.000	0.525	0.367
P_{RL}	1.000	0.475	0.291

1) The 1997 HCM, Table 9-4, p 9-12.

4.1.2 Additional Portion of Queue Service Time

The original NCHRP method considers the length of an actuated phase as two separated portions, queue service time (G_s) and green extension time (G_e) [1, 7, 18]. The method estimates the average lengths of those two periods individually and the length of a whole phase by summing those two. G_s is the green time period that vehicle actuations occur sequentially with the gap that is less than unit-extension time due to the queued vehicles. The original NCHRP method suggests that it is equivalent to the end point of the queue-accumulated polygon. However, it should be larger than that since the back-of-queue locates much farther than the end of queue-accumulated polygon on the upstream of the approach. The time taken with the last vehicles to drive from the end of back of queue to the stop line should be considered as a portion of G_s. Figure 4-3 presents the time-space diagram.

The additional time (G_a) is newly introduced. It is the time taken from the time that the last vehicle from the back of queue starts to move to the time that vehicle leaves detection area on an approach. It was estimated by the following equation:
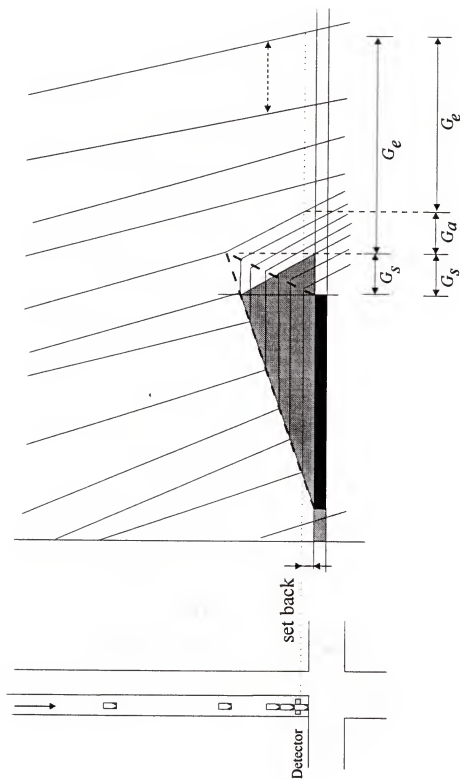
Figure 4-3: Additional portion of queue service time

$$G_a = \left(d_L Q_B - d_s\right)/(1.47s) \tag{4-5}$$

where

d_L = head-to-head vehicle gap distance in queue,

Q_B = back of queue (ft),

d_s = detector set back distance (ft) and

s = average cruising speed (mph).

The G_a was estimated based on the length of back of queue, average cruising speed and the average distance from head to head of vehicles in the queue, 18 ft. It was added into G_s in the time-step based method.

4.2 Other Improvements

Other improvements mainly include the replacement of the shared-permitted left-turn-parameter estimation models used in the original NCHRP method with (1) the ones recently introduced in the field of study and (2) the adjustment of the modeling approach. They are summarized in the following list:

1. The original NCHRP method assumes that the free green (g_f), a portion of green time that through vehicles can flow on a shared left-turn lane before the first arrival of the permitted left-turn vehicle, should always be less than the queued green (g_q), a portion of green time that permitted left turns cannot flow due to the queued opposing through vehicles flowing into the intersection. In the modified procedure, it was allowed that g_f could be longer than g_q. With the improvement, it is expected that the

overestimation of phase length due to the misestimated arrival of the first shared left-turn vehicle be prevented.

2. The original NCHRP method employs the shared-permitted left-turn-parameter estimation models suggested in the 1994 HCM. The models were replaced with the ones from the 1997 HCM. The 1997 HCM provides the improved models, and they remain in the 2000 HCM. However, it should be noted that the modified NCHRP method still employs the 1985 HCM model for the estimation of percent of left turns on a shared lane (P_L) since the P_L estimation models from both the 1994 and 1997 HCM provide oscillation of phase length over iterations.

3. For the estimation of the number of pedestrians arrived on red, it was assumed in the original NCHRP method that the effective red times of the vehicles affect pedestrian arrival. It was updated that the number of pedestrians arrived on red is only a function of displayed red time but not of the effective red time. With the modification, it was expected to minimize the estimation error anticipated with the compound left-turn treatment, where effective red time is significantly less than the displayed red time.

4. The original NCHRP method failed to count the opposing shared left-turn volume as the opposing volume. Improvements were made to count the shared left-turn volume on opposing approach as the opposing volume as defined in the HCM. The modification was intended to prevent the

underestimation of the phase length due to the miscounted opposing volume of the permitted left turns.

5. With the simple permitted left-turn operation, the arrival rate of the permitted left turns was increased in the original NCHRP method by multiplying a vehicle-equivalent factor for left turns, EL_1. The volume modification was eliminated to prevent the overestimation of the average phase length since the adjusted saturation flow rate from HCS was used.

6. The original NCHRP method considers the phase-skipping effect in the estimation of the average queue service time and the green extension time. As a modification, the phase-skipping effect is not considered in the estimation of the queue service time. This is because the queue service time estimation model employs the assumption of uniform arrival. The phase-skipping effect should only be considered in the estimation of the green extension time, where bunched exponential headway distribution is assumed.

7. The original NCHRP method uniformly treats the compound left-turn cases. Compound left turns should be considered separately based on the sequence of the phases: permitted plus protected and protected plus permitted. Those two cases were handled as two independent phasing operations in the updated NCHRP method since the left-turn queue at the beginning of the permitted period would be different in those cases.

The NCHRP method with the modification described above was coded with the Visual Basic computer programming language for use in the maximum green time

parameter design procedure. The performance of the modified NCHRP method was compared to the one of the original NCHRP method based on the CORSIM simulation data, surrogating field data.

4.3 Test of the Modified NCHRP Method

The modified NCHRP method was coded with the Visual Basic computer programming language, and the average phase lengths estimated by the modified and original NCHRP method were compared to demonstrate the improvement. A set of hypothetical intersections was developed to consider various geometric conditions, as specified by the shared permitted worksheets in the HCM. The CORSIM simulation data surrogating field data were used in the test due to the limitations of field data since the data should reflect many possible combinations of traffic and control conditions. The CORSIM simulation data are not statistically different from the field data and are widely accepted as subrogation of field data [29].

Four different intersections were designed. The only lanes on the east-westbound approaches were varied, while the north-southbound approaches had a single lane in order to minimize the number of test data determined by the combination of phasing sequence and operation. Intersection 1 presented in Figure 4-4 covers the shared permitted left turns opposed by a single lane approach. Intersection 2 presented in Figure 4-5 covers the shared permitted left turns opposed by multiple lanes including a permitted left-turn lane. Intersection 3 presented in Figure 4-6 covers the case of shared permitted left turns opposed by multiple lanes with a protected left-turn lane. Intersection 4 presented in Figure 4-7 covers the case where a protected left-turn lane is provided on both opposing approaches.
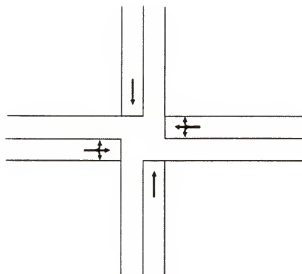
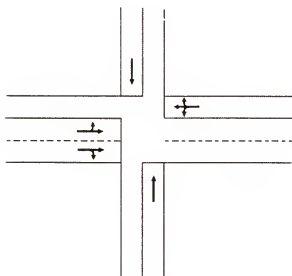Figure 4-4: Hypothetical Intersection 1

Figure 4-5: Hypothetical Intersection 2
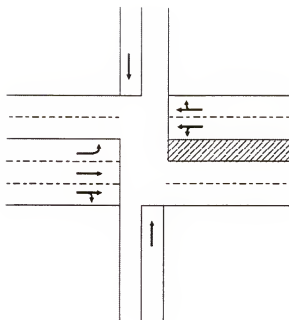
Figure 4-6: Hypothetical Intersection 3
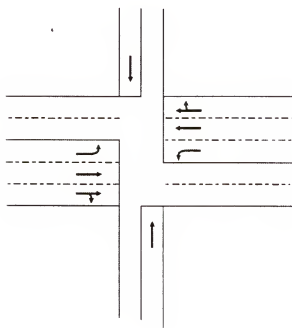
Figure 4-7: Hypothetical Intersection 4

Sets of phasing sequences for traffic-actuated operations were designed to consider various control conditions for the test. A set of phasing sequences was designed for each hypothetical intersection. They are presented in Tables 4-2, 4-3 and 4-4. A simple two-phase operation presented in Table 4-2 was considered for the operation at the Intersections 1 and 2. Five combinations of permitted and protected left-turn treatments and leading and lagging sequences were considered for Intersection 3. The eleven sets of permitted and protected left-turn treatments and leading and lagging sequences were considered for Intersection 4.

Table 4-4: Phase operation designed for Intersections 1 and 2

Phase number	1	2
Movement flowing		

Table 4-5: Phase operation designed for Intersection 3

Case Index	East-westbound left-turn treatment		Phase number		
	Permitted/Protected	Leading/Lagging	1	2	3
	Permitted	-			
	Protected	Leading			
		Lagging			
	Permitted/Protected	Leading			
		Lagging			

Table 4-6: Phase operation designed for Intersection 4

Case Index	East-westbound left-turn treatment		Phase number			
	Permitted/Protected	Leading/Lagging	1	2	3	4
	Both Permitted	-				
	EB Protected & WB Permitted	Leading				
		Lagging				
	Both EB & WB Protected	Leading				
		Leading/Lagging				
	EB Permitted/Protected WB Permitted	Leading				
		Lagging				
	EB Permitted/Protected WB Protected	Leading				
		Leading/Lagging				
	Both EB & WB Permitted/Protected	Leading				
		Leading/Lagging				

It should be noted that the left-turn phase sequences with all lagging protected left turns were excluded from the sets since they are out of bounds of the NCHRP method. In all intersection cases, eastbound was controlled as the major direction by assigning higher traffic volume than the opposing approach in the test. Thus, the phase sequence that may be observed when westbound traffic volume is higher than eastbound was excluded from the phase sequence representation.

To take account of the various traffic conditions in the test, a basic traffic-volume condition for each intersection was initially set, and then volume variations were made by applying adjustment factors to the condition. The basic traffic condition for each intersection case was determined by the following procedure. First, a link volume was determined, based on 500 vehicle-per-hour-per-lane (vphpl) for a through lane, 200 vphpl

for an exclusive turning lane and 100 vphpl for a shared turning lane. Second, approach volume was set with 60:40 and 50:50 directional balances for east-westbound and north-southbound streets, respectively. Since the eastbound approach was taken as a major movement, for the east-westbound street, 60 percent of the link volume was assigned to the eastbound approach, and the rest was assigned to the westbound approach. Third, fifteen percent of the approach volume was set for each turning movement, and the rest was given to the through movement.

Based on the standard condition, a set of different traffic conditions was generated by providing variations in total approach volume, approach balance and percent left turns. The specific numbers used for the volume variations are summarized in Table 4-5. Four different levels of total approach volumes were considered by multiplying 0.5, 0.75, 1.25 and 1.5 to the total approach volume of the basic condition. Four different levels of approach balances were considered, based on the link volume specified in the basic condition: 50:50, 55:45, 65:35 and 70:30. Also, four different levels of percent left turn were considered. The percent left turn was changed to 5, 10, 20 and 25 percent from the approach volume specified in the basic condition.

Table 4-7: Variation applied to the standard condition to change the traffic condition

Items	1	2	3	4
Total traffic volume (times)	0.50	0.75	1.25	1.50
Approach balance	50:45	55:45	65:35	70:30
Left-turn balance (% of total volume)	0.05	0.10	0.20	0.25

The combination of the hypothetical intersections (geometric conditions), phase sequences (control conditions) and volume variations (traffic conditions) introduced 234 different cases of intersection operations. They introduced 1196 different phase lengths to compare throughout the test.

4.3.1 Test Procedure

The original NCHRP method was implemented by using the ACT3-48 program, developed by Lin at the University of Florida in 1996 as the final product of the NCHRP 3-48 project. The average phase lengths estimated by the original and the modified NCHRP methods were compared to the ones counted from the CORSIM simulation results. Since CORSIM does not provide average phase length as an output in its text output file, the signal timing data stored in the TSD file were read. TSD is a binary file, generated by CORSIM for the TRAFVU animation. The CORSIM average phase lengths were calculated based on the TSD data. A program called TSDREADER was developed to assist the CORSIM signal timing reading procedure by utilizing CORTOOL, the ActiveX tool that provides an access to the TSD simulation data [41]. TSDREADER reads the TSD signal timing data and computes the average phase lengths over different cycles based on those.

In the simulation study, multiple CORSIM runs were made with a different set of random seed numbers for each simulation case. First, 10 initial analyses were made. When the standard deviation of the 10 different average phase lengths is less than or equal to the 10 percent of the overall average of the 10 different average phase lengths, the simulation process was terminated. If not, the sample size, the number of CORSIM runs, was controlled to be increased by adding 10 more CORSIM simulation runs with newly introduced random seed numbers. However, it turned out that 10 initial runs were

enough to satisfy the sample-size-control strategy used; all standard deviations were less than or equal to 10 percent of the overall average.

It is critical to set the CORSIM simulation parameters as close as possible to follow the assumptions embedded in the analytical models used in the NCHRP method. Three assumptions found in the NCHRP method were controlled to be followed in the simulation as well. They include 4.5 seconds of critical gap of permitted left turns, no left-turn jumper and two left-turn sneakers. In the CORSIM runs, those were controlled by providing record types 145, 140 and 141 into the TRF input file. With record type 145, 4.5 seconds of critical gaps were uniformly applied for 10 different CORSIM driver types. With record type 140, no left-turn jumper was set. With record type 141, two left-turn sneakers were set by setting 100 and zero percent probabilities of left-turn moving within four seconds after the end of green time and after four seconds after the end of green time, respectively. Since the saturated headway is two seconds, two jumpers would be allowed to flow for four seconds.

As a result, three sets of the estimated average phase lengths were obtained. The models implemented include the modified NCHRP method, the original NCHRP method and CORSIM. Each set of average phase lengths estimated was divided into two cases: the protected exclusive lane and the shared permitted left-turn cases. When a shared permitted left-turn lane is provided, a detector should be installed on the lane to detect both through and left-turn vehicles. Thus, the actuations due to left-turn vehicles stopping at the stop line and waiting for an acceptable gap from the opposing traffic affects the extension of green extension time. This case was separately compared. The

case of the permitted left turns on an exclusive lane was out of consideration since no actuation is required on that lane.

4.3.2 Performance with Protected Movements on Exclusive Lanes

The 1,118 phases operating with the protected through and the exclusive protected left-turn movements were counted in this group. They included most of the test phases but excluded only two: the eastbound through phase at Intersection 2 and the westbound thorough phase at Intersection 3. The average phase lengths were estimated with 10 sec of minimum green, 46 sec of maximum green and three seconds of unit extension time. The approaching speeds on major and minor streets were set to 40 and 30 miles per hour (mph).

The average phase lengths estimated by three different models were graphically compared. Figure 4-8 presents the comparison made between the average phase lengths estimated by CORSIM and the original NCHRP method.
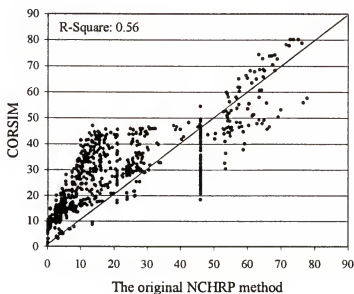
Figure 4-8: Average green time comparison between CORSIM and the original NCHRP method

The average phase lengths that are longer than the maximum green time, 46 sec, are the ones extended for the simultaneous termination with its concurrent phases. The results illustrate that the existing NCHRP3-48 method underestimates the average green time before it reaches to its maximum green time, 46 sec. It is also shown that the average phase lengths estimated by the NCHRP method reaches its maximum so rapidly that a barrier appears at 46 sec on horizontal axle. The underestimation from the original NCHRP method would partially be because of the saturation flow rate briefly estimated by the WHICH program, a data input manager program assisting the execution of ACT3-48. Mainly, it is because of (1) the phase-skipping effects applied to both green extension and queue service times and (2) the definition of queue service time used in the original method.

On the other hand, Figure 4-9 presents the comparison between the modified NCHRP method and CORSIM. The results show that the modified NCHRP method provides more of a one-to-one relationship to the average phase lengths obtained through the CORSIM simulations than the original NCHRP method does. The modified NCHRP method uses the saturation flow rates directly estimated with the HCM method. The comparison chart demonstrates the improvement introduced with the updated method.

The R^2 values between the phase lengths estimated by each of two models (the original and the modified NCHRP methods) and CORSIM are 0.56 and 0.90, respectively. It was found from those two comparison charts and the R^2 values that the modified NCHRP provides much more reliable estimation than the original NCHRP method does, at least for the case of exclusive protected operation.
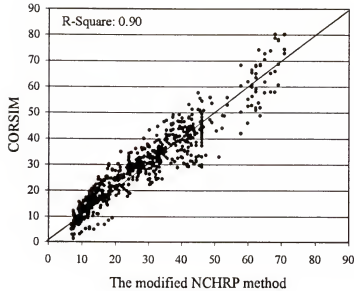
Figure 4-9: Average green time comparison between CORSIM and the modified
NCHRP method

4.3.3 Performance with Shared Permitted Left Turns

The 78 average lengths of phases that allow shared permitted left-turn movements
to flow on a multiple-lane approach were separately compared. The case includes two
cases of the test phases: the eastbound through phase at Intersection 2 and the westbound
through phase at Intersection 3. The average phase lengths were estimated with 10 sec of
minimum green, 46 sec of maximum green and 3 sec of unit extension time. The
approaching speeds on the major and minor streets were set to 40 and 30 miles per hour
(mph). Figure 4-10 illustrates the first comparison.

It was also shown that the original NCHRP method overestimated the average
phase length in general. With low traffic volume, the method underestimated the average
phase lengths due to the phase-skipping effects applied to the whole phase length. When
traffic volume is high, the original NCHRP method overestimates the average phase
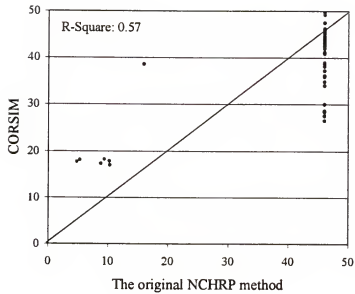
Figure 4-10: Average green time estimated by the NCHRP method for the permitted phases and the phases affected by permitted movements

lengths due to the queue service time being increased by permitted left turns, and most of the phases reached the maximum green time of 46 sec. Permitted left turns wait on the green for acceptable gaps in the opposing traffic stream at the stop line. Estimation of the proper length of the queue service time leads to the proper estimation of average phase length, and the estimation of the proper percentage of left turns on the shared lane leads to the proper estimation of queue service time.

The comparison made between the modified NCHRP method and CORSIM is illustrated in Figure 4-11. The Figure shows that the modified NCHRP method provides more close one-to-one relationships to the CORSIM simulation results than the original NCHRP method does. However, the overestimation problem found with the original NCHRP method has been observed with the modified NCHRP method.
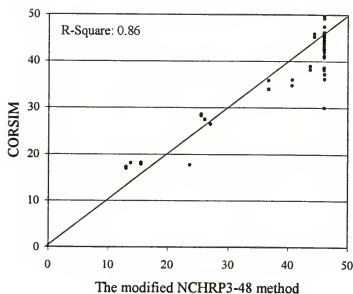
Figure 4-11: Average green time estimated by the modified NCHRP method for the permitted phases and the phases affected by permitted movements

Figures 4-10 and 4-11 demonstrate that the modified NCHRP method provides more close one-to-one relationships to the CORSIM results than the original NCHRP method does for the case of shared-permitted-left turns. The overestimation problem observed in both the original and the updated NCHRP methods still remain. It should be noted that the overestimation tendency is a problem not of the NCHRP method but of the shared left-turn parameter estimation models suggested in the HCM. Modification in the shared permitted left-turn parameter estimation model is required.

4.4 Effect of Maximum Green Parameters to the Average Phase Length

The comparison tests conducted in the previous sections are for a single value of the maximum green parameter, 46 sec. A new test that verifies the marginal effect of the maximum green parameter to the estimated average phase length was made and presented

prior to the maximum green time optimization. With a simple hypothetical intersection and a fixed traffic condition, a set of average phase lengths was estimated based on a set of the maximum green time parameter values. The changing trend of the average phase length was compared with the one observed with the CORSIM simulation.

4.4.1 Average Green Time vs. Maximum Green Time

A new hypothetical intersection was adopted for the study. The hypothetical intersection presented in the final report of the NCHRP 3-48 project was adopted for the test. The intersection is presented in Figure 4-12. It is configured with four single-lane approaches, and each approach has 675 vph demand. The intersection is controlled with a simple two-phase sequence. The actuated-control parameters, the minimum, maximum and unit extension times, were set to 10, 46 and 3 sec, respectively. By increasing the maximum green time parameters from 15 to 100 by 5 seconds, 18 different average phase lengths were obtained with the modified NCHRP method and from the CORSIM simulation.

Average headway = 2.0 seconds
Lost time per phase = 3.0 seconds
Detector length = 30 ft
No setback from the stop line
Approach speed = 30 mph

$V^* = 675$ vph

$V^* = 675$ vph

	Phase 1	Phase 2
Gmin	10	10
Unit extension	3	3
Gmax	46	46
Intergreen	4	4
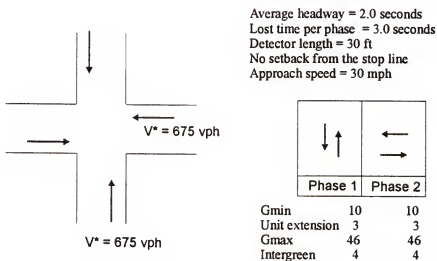
Figure 4-12: The hypothetical intersection used for the test of the NCHRP method

In the simulation study, again, multiple CORSIM runs were made with different random seed numbers for each simulation case. First, 10 initial analyses were made. When the standard deviation of the 10 different average phase lengths was less than or equal to the 10 percent of the overall average of the 10 different average phase lengths, the simulation process was terminated. If not, the sample size, the number of CORSIM runs, was controlled to be increased by adding 10 more CORSIM simulation runs with newly introduced random-seed numbers. The average phase lengths estimated by the modified NCHRP method and CORSIM were obtained. The changing trends of the average phase lengths over various maximum green-time parameters from those are graphically compared and presented in Figure 4-13.
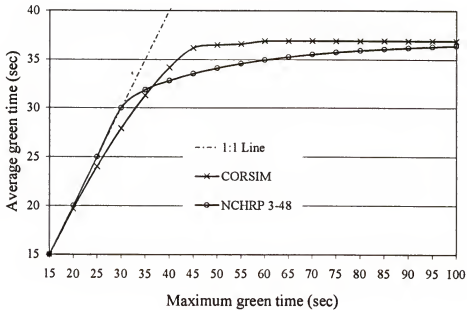
Figure 4-13: The maximum green-time parameters vs. the average green times

The figure shows that the modified NCHRP method is capable of explaining the changing trend of the average green over the maximum green. The result shows that,

when the maximum green time parameter is long, the average green time estimated by CORSIM is a little longer than the one estimated by the NCHRP method. However, the largest gap observed in the average greens from those two methods is 2.65 sec, which is less than a unit extension time, 3 sec, at 45 maximum green. In short, the graph shows a similar changing trend of average green time over the maximum green parameter since less than one green extension is observed in the extreme case.

The followings were found from the results. First, a close one-to-one relationship between the maximum green and the average green was observed when short maximum green parameters were provided. This implies that the green time is fully used and gapout rarely occurs. Second, the maximum green time parameter that is longer than the average green time plays an insignificant role in traffic-actuated operation. Gapout always occurs at a fixed point in the middle of green before the green reaches the maximum green. This implies that providing extremely long maximum green times result in the same average green time, no matter how long the parameters are.

4.4.2 Design of Maximum Green Time Parameters

The relationship between maximum green parameters and the average green time shows that it would be hard to determine the optimal point when maximum green parameters become large since the average green time would not be significantly varied. This implies a nonconvex delay changing trend and no minimum point. Therefore, an additional test was made to verify the delay-changing trend in three-dimensional space with axles of a maximum green parameter, the average green time and the average red time, which is a function of the average green time of the other phases. The hypothetical intersection presented in Figure 4-12 was used for the test. Again, the maximum green time parameter was varied from 15 to 100 seconds by increasing 5 seconds for both

phases. A total of 324 (18×18) combinations of maximum green time parameters were tested and plotted. For each set of maximum green time parameters, four different levels of traffic volume were tested. They were 675, 780, 820 and 900 vphpl.

Four paired charts are presented in Figures 4-14 to 4-21 for four different volume levels. They are plotted based on the results obtained from the modified NCHRP method and the HCM-based evaluation. The modified NCHRP method was combined with the proposed HCM-based signal timing design procedure presented in the previous chapter in order to design the maximum green time parameters. In the design procedure (the proposed HGA searching), the modified NCHRP method is implemented prior to the HCM evaluation to obtain the average phase lengths. Then, the HCM evaluation is executed (the average phase length estimated by the modified NCHRP method was evaluated by treating those as the pretimed phases), and the objective function value is computed. The best set of the maximum green times that yields the best set of average phase lengths satisfying the designed optimization strategy (aggregated delay minimization) was obtained at the final HGA iteration. The three-dimensional chart was plotted to demonstrate the pseudo-convex shape of the objective function, the overall average delay, and the two-dimensional chart was plotted to present the minimal point of the function.

Figures 4-14 and 4-15 demonstrate that, when large values of maximum green time parameters are provided, the overall delay increases. This implies that the average green times be truly changed when the maximum green time parameters are changed. This trend was not observed in the previous test performed with a single phase (see Figure 4-13). The proposed maximum green time parameter design method, the
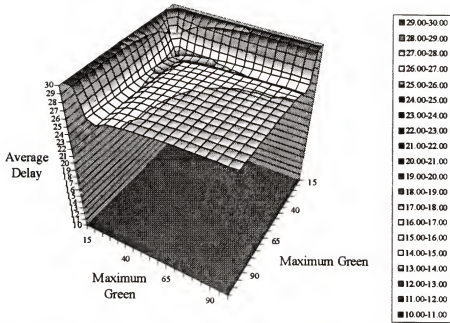
Figure 4-14: Delay changing trend (3D) over maximum green times based on the HCM and the modified NCHRP methods with 675 vphpl
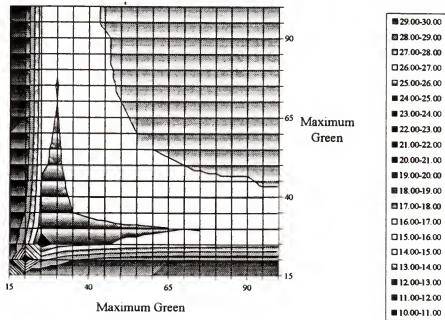
Figure 4-15: Delay changing trend (2D) over maximum green times based on the HCM and the modified NCHRP methods with 675 vphpl
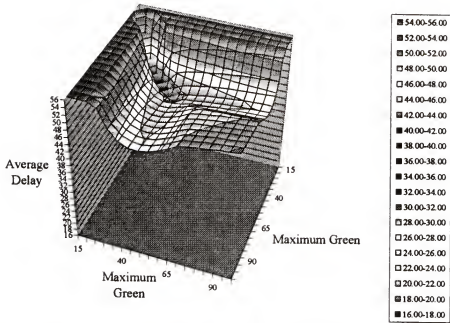
Figure 4-16: Delay changing trend (3D) over maximum green times based on the HCM and the modified NCHRP methods with 780 vphpl
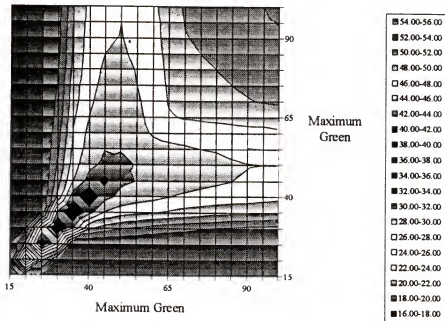
Figure 4-17: Delay changing trend (2D) over maximum green times based on the HCM and the modified NCHRP methods with 780 vphpl
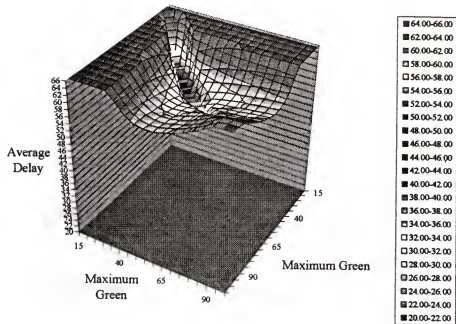
Figure 4-18: Delay changing trend (3D) over maximum green times based on the HCM and the modified NCHRP methods with 820 vphpl
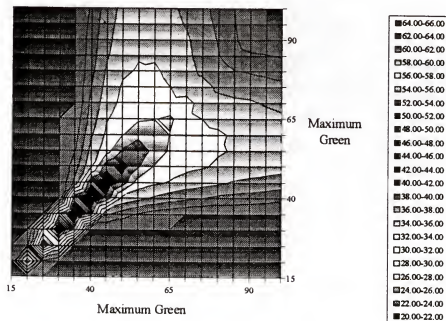
Figure 4-19: Delay changing trend (2D) over maximum green times based on the HCM and the modified NCHRP methods with 820 vphpl
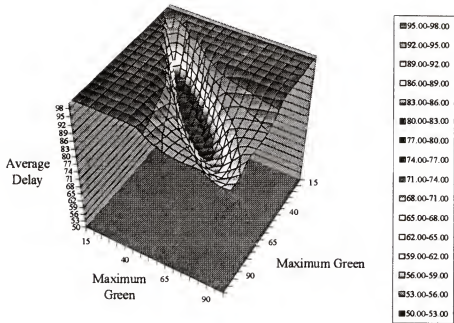
Figure 4-20: Delay changing trend (3D) over maximum green times based on the HCM and the modified NCHRP methods with 900 vphpl
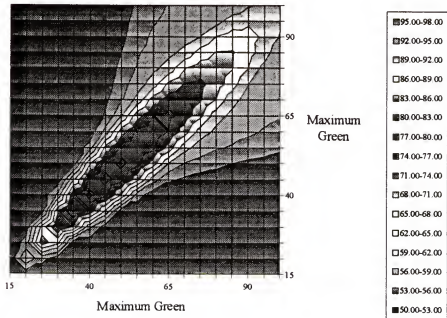
Figure 4-21: Delay changing trend (2D) over maximum green times based on the HCM and the modified NCHRP methods with 900 vphpl

combination of the modified NCHRP method and the HGA computational procedure presented in Chapter 3, was applied with the same traffic, geometric and control conditions specified above. It has been found that the proposed HGA computational method is capable of finding the minimal points of those pseudo-convex functions. For four different input volume levels, the best values found by the proposed method were 20, 30, 35 and 40 sec, respectively. The tests demonstrate that the proposed method, combining the modified NCHRP method and the HCM-based evaluation, is capable of finding the optimal set of maximum green time parameters in terms of the HCM optimum.

CHAPTER 5
TESTING AND EVALUATION

Previous chapters of this dissertation have proposed new algorithms and modifications to existing algorithms that have the potential to advance the state of the practice in signalized intersection timing design and analysis. Chapter 3 presented a new optimization scheme based on a combination of hill climbing and genetic algorithms. Chapter 4 proposed a set of modifications to an existing model for the estimation of phase times at traffic-actuated signals. This chapter describes the testing and evaluation of the enhancements proposed in the previous chapters.

The testing involves comparison of the performance of the proposed HCM-based signal timing plans developed by the enhanced models with that of the plans designed by the existing (non HCM-based) software products. This chapter presents the procedures undertaken in the comparison tests along with their results. A version of the Signal Operation Analysis Package referred to as SOAP2K has been developed for purposes of these tests. The evolution of SOAP as a traffic-engineering tool was originally described in the literature review in Chapter 2. SOAP2K incorporates both the proposed hybrid genetic algorithm from Chapter 3 and the modified NCHRP method [42] from Chapter 4.

In addition, the chapter presents a description of an XML-based scheme that stores the specific intersection information for bi-directional transfer between SOAP2K and the HCS. This scheme is referred to as the Traffic Model Markup Language (TMML).

The pretimed signal timing optimization procedures were tested for both signal timing design strategies selected for this study: (1) average delay minimization to all vehicles and (2) control delay equalization among all critical lane groups. For average delay minimization, the signal timing designed by the proposed method was compared with the timing from TRANSYT-7F. For maximum green time optimization the maximum green time parameters designed by the proposed method were compared with those from existing methods described in Chapter 2. Lacking the resources to conduct a major field study to obtain empirical comparison data, the performance evaluations produced by the CORSIM simulation model (also described in Chapter 2) were used as a surrogate for field data.

5.1 Implementation of the Proposed Computational Procedure

SOAP2K has been developed for the efficient implementation of the proposed computational procedure. It contains the functionality to estimate the average length of an actuated phase based on the modified NCHRP method and the ability to communicate with HCS. The computer-based computation structure of SOAP2K designed for the automation of the test procedure is presented in following subsections.

5.1.1 Computational Structure

The structure of the proposed HCM-based signal timing design procedure consists of two major computer programs: SOAP2K and HCS. SOAP2K was complied as a dynamic linked library that can be called from its administrative program, HCS. HCS is the program developed for the automation of the HCM evaluation procedure at the University of Florida under the supervision of FHWA [30]. The programmatic structure of the proposed computational procedure is illustrated in Figure 5-1.
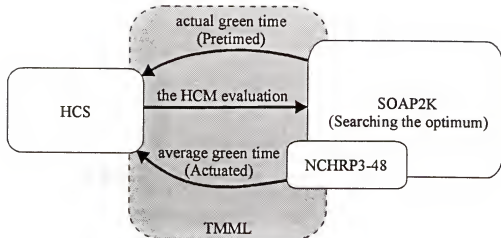
Figure 5-1: The structure of the proposed computational procedure

For pretimed signal timing design, SOAP2K (the proposed HGA computational method) determines the values decision variables (green times) at the initial stage and during the searching procedure. During the searching procedure, the signal timings determined by SOAP2K are evaluated by HCS. Then, the objective function value is obtained and the next searching point is determined based on the value of the objective function by SOAP2K. The HGA search continues until it reaches the optimum for pretimed operation. The Traffic Model Markup Language (TMML) was developed and used for communication between those two programs. The Appendix provides the description of TMML.

For maximum green time design, the proposed HGA computational method deals with the maximum green time parameters instead of the green times as decision variables. The searching structure is fundamentally similar to the one described above. The differences are (1) that the modified NCHRP method estimates the average green times of the actuated phases and (2) that HCS evaluates the signal timing based on the

estimated average green times. The proposed HGA computational method (SOAP2K) and the HCM-based evaluation method (HCS) keep performing such iteration until they reach the best set of maximum green time parameters for traffic actuated operation.

5.1.2 Development of SOAP2K

SOAP2K utilizes (1) the input screen of HCS for general data and (2) its own input interface for such traffic-actuated control parameters as minimum green and maximum green times, which are not required for the HCM evaluation. The SOAP2K input interface and the SOAP2K timing design screen are presented in Figures 5-2 and 5-3. SOAP2K and HCS share the four-bit memory to communicate with each other. Through the memory, a signal flag indicating the status of computation controls the order of the execution of those programs. The signal flag works like a baton providing the right of execution. For example, when SOAP2K gives the signal to HCS, it stops

Figure 5-2: Input screen of SOAP2K

Figure 5-3: Timing design screen of SOAP2K

computation and waits for the signal back from HCS. When HCS receives the signal, it starts to execute the HCM evaluation, and then it provides the signal back to SOAP2K after the evaluation. SOAP2K and HCS share the data through TMML. Whenever one of those programs terminates its computation during the iteration, the TMML data file is generated in order to share the data with the other. TMML is the universal database developed during the research.

5.2 Test Methodology

A set of hypothetical intersections was used in the comparison test performed for the strategies of average delay minimization and control delay equalization among critical lane groups. With the various traffic and control conditions specified based on the hypothetical intersection data, the signal timings were designed by the proposed HGA

computational procedure and two different versions of TRANSYT-7F based on those strategies. The CORSIM simulation data surrogating the field data were used in the test due to limitations in the field data collection. The signal timings designed by the selected methods were compared based on the CORSIM simulation results.

5.2.1 Test Data

Six hypothetical intersections were set to include different geometric conditions in the test. They are the four intersections used in the test presented in the previous chapter (see Figures from 4-3 to 4-6) and two T-intersections. The four intersections cover the combination of a single-lane/multilane opposed to a single-lane/multilane approach at the four-approach intersection. The two T-intersections used in the test cover one-way and two-way operation at T intersections; one features with the ordinary two-way operation, and the other features with one-way operation. They are presented in Figures 5-4 and 5-5.

Various traffic conditions were generated based on the rule used in the test of the modified NCHRP method. The rules used to generate various traffic conditions,
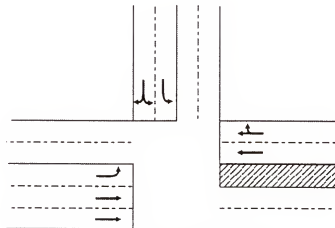
Figure 5-4: Hypothetical intersection 5
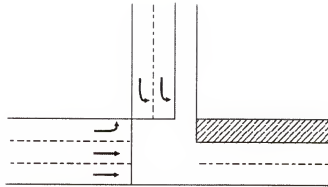
Figure 5-5: Hypothetical intersection 6

including how to determine the basic traffic condition, are described in the previous chapter. The identical values of adjustment factors used in the previous chapter are also applied. For Intersection 6, the approach balance adjustment was not applicable due to the one-way operation. Only approach volume and left-turn percent were adjusted for Intersection 6.

Various signalization conditions were made, based on the possible combination of phase operations and sequences. They reflect the combination of simple permitted, simple protected and compound left turns. The concurrent lagging protected left-turn cases that were excluded in the test of the modified NCHRP method were included in this case. Therefore, the possible number of phase operations for Intersection 4 changes to fourteen. A new set of phase operations for the intersection is presented in Table 5-1. The combination of the phase operation and sequence for Intersection 5 and 6 are presented in Tables 5-2 and 5-3. The whole combination of traffic, geometric and control conditions yields 347 different intersection analysis cases.

Table 5-8: Phase operation designed for Intersection 4

Case Index	East-westbound left-turn treatment		Phase number			
	Permitted/Protected	Leading/Lagging	1	2	3	4
	Both Permitted	-				
	EB Protected & WB Permitted	Leading				
		Lagging				
	Both EB & WB Protected	Leading				
		Lagging				
		Leading/Lagging				
	EB Permitted/Protected WB Permitted	Leading				
		Lagging				
	EB Permitted/Protected WB Protected	Leading				
		Lagging				
		Leading/Lagging				
	Both EB & WB Permitted/Protected	Leading				
		Lagging				
		Leading/Lagging				

Table 5-9: Phase operation designed for Intersection 5

Case Index	East-westbound left-turn treatment		Phase number		
	Permitted/Protected	Leading/Lagging	1	2	3
	Permitted	-			
	Protected	Leading			
		Lagging			
	Permitted/Protected	Leading			
		Lagging			

Table 5-10: Phase operation designed for Intersection 6

Phase number	1	2
Movement flowing		

In the multiple CORSIM simulation analysis, the rules used in the test of the modified NCHRP method were applied, and they were described in the previous chapter. The number of the CORSIM runs for a single analysis case was determined if the standard deviation of delays was less than 10 percent of the mean value. The 10 initial runs were made with the different random seed numbers. If the standard deviation was less than or equal to 10 percent of the average, no more runs are required. If it is larger than 10 percent of the averages, the sample size has been increased by adding 10 more CORSIM runs with 10 more sets of different random seed numbers.

A program named CORAST was coded with the Visual Basic programming language. CORAST reads the signal timing data from the SOAP2K and TRANSYT-7F outputs and generates a set of TRF files, CORSIM input files, with different random seed numbers predefined. It controls the number of CORSIM runs based on the simulation sample size determination strategy described.

In the CORSIM analysis, the critical gap of permitted left turns, left-turn jumpers and left-turn sneakers were controlled with record types 145, 140 and 141, respectively. Four point five seconds of critical gap were constantly used for all types of drivers in the record type 145. No left-turn jumper was set as assumed in the HCM through record type 140. Up to two left-turn sneakers were allowed to precede their maneuver per cycle

through record type 141 by setting 100 percent probability of moving if they arrive at their stop line within four seconds after the end of the green time.

5.2.2 Test Description

Computers with Pentium II processors, 400 Mhz speed, were used for the test. The TRANSYT-7F runs were made by coding and implementing a program named TRAP. TRAP generates a set of TIN files, the TRANSYT-7F input files, and executes the TRANSYT-7F program. Then, CORAST was applied to conduct CORSIM analyses. Two versions of TRANSYT-7F were applied: versions 7.0 and 9.2. The 8.x version of TRANSYT-7F was not able to handle the delay-only minimization strategy by setting a zero-stop penalty due to the problem embedded in the program. Version 7.0 was tested at the initial stage of the test. When the new version of the program, 9.2, became available at the final stage of the study, the version of the program was included in the test.

The multiple CORSIM runs were assisted with the RUNCOR program utilizing CORSIM.DLL. Instead of using the latest version of CORSIM, version 5.0 beta, the CORSIM version 4.32 was employed for the test. It has been experienced with several test data sets that the CORSIM version 4.32 and HCS provide similar control delays, while the CORSIM version 5.0 beta provides incomparable delay.

5.3 Signal Timings for Pretimed Control

For the average delay minimization strategy, the control delays estimated by CORSIM with the signal timings designed by the proposed HGA computational procedure and TRANSYT-7F were compared. For the strategy of control-delay equalization among all critical lane groups, the highest HCM delays of the critical lane groups on major and minor streets were compared. The following subsections describe the results.

5.3.1 Average Delay Minimization

The overall intersection delays estimated based on two different sets of signal timings designed by the proposed HGA computational method and TRANSYT-7F were compared through CORISM. The comparison results were plotted and are presented in Figure 5-6.
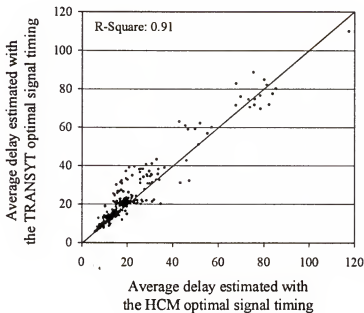
Figure 5-6: Delay comparison between TRANSYT-7F version 7.0 and the proposed computational procedure

The figure illustrates that the plots are biased to the upper side of the one-to-one line in the chart. This shows that the signal timings designed by the HCM signal timing design procedure provide less delay than the ones from TRANSYT-7F. This is because the TRANSYT-7F uses a roughly estimated saturation-flow rate. In other words, this is because the proposed HGA computational procedure updates and uses the HCM adjusted saturation flow rate for permitted left turns during the searching. The saturation flow rate of permitted left turns should be adjusted based on the changed signal timings given for

the movement and its opposing movement. TRANSYT-7F uses a fixed saturation flow rate for the permitted movement without consideration of signal timings changing in its searching procedure. This weakness of TRANSYT-7F provides a tendency for higher delay than the one provided by the HCM.

The delay estimated with the signal timings designed by TRANSYT-7F version 9.2 was compared to the ones from the proposed computational procedure. Figure 5-7 shows the comparison results.
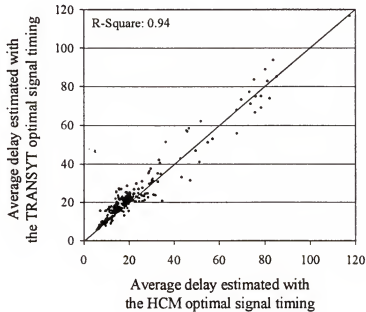
Figure 5-7: Delay comparison between TRANSYT-7F version 9.2 and the proposed computational procedure

The figure shows that the plots are still slightly biased to the left side of the one-to-one line. The improvement made in TRANSYT-7F version 9.2 introduces better results than TRANSYT-7F version 7.0. The major improvement includes the changes in simulation method; version 9.2 uses step-wise simulation, while version 7.0 uses link-

wise simulation [43]. With the step-wise simulation used in version 9.2, enhancement was provided to the permitted left-turn performance analysis since link-wise simulation is unable to model the complex nature of permitted movements properly.

The HCM uses an analytical approach to evaluate the performance of permitted left turns, while TRANSYT-7F uses the step-wise simulation. In order to make graphical comparisons for the effect of those different analysis approaches for permitted left turns, the test results from the comparison between TRANSYT-7F version 9.2 and the proposed computational procedure were categorized into two groups. As a group, the cases of the intersections operated with the protected phases only were categorized. The cases of the intersections operated with at least one permitted left turn involved were categorized into the other. The first group contains the results from the protected phases only. All movements were controlled by the protected phases. The second group contains the results from both protected and permitted phases; the protected phase lengths designed by being affected by permitted phases in terms of red time were categorized into the second group. The first group contains (1) Intersection 4 with phase operations four, five and six, (2) Intersection 5 with phase operations two and three and (3) Intersection 6 with phase operation one. A total of 74 cases were categorized as a group. The 273 cases were categorized as the other.

Comparison filtered for group one is presented in Figure 5-8. The figure illustrates that the signal timing designed by the HCM-based design method, the proposed HGA method, resulted in less delay; the plots appeared in the left side of the one-to-one line. For protected phases, it was shown that the phase lengths designed by

the proposed computational procedure provided the trend of lesser delay more than the other.
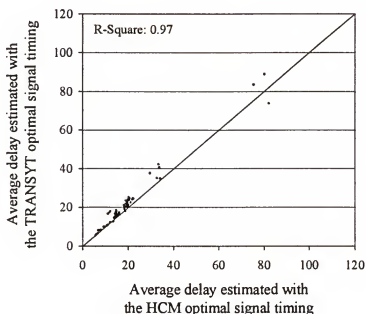
Figure 5-8: Delay comparison between TRANSYT-7F and the proposed method for protected phases

Figure 5-9 shows the comparison made for the permitted left-turn involved operation, affected by the combination of protected and permitted phase operation. The figure illustrates that the significant amount of plots are still the left side of the one-to-one line. However, it was observed that TRANSYT-7F provides better signal timings than the HCM-based method in some cases. This is because of the weakness of the HCM evaluation procedure for permitted left turns.

The HCM uses a straightforward worksheet procedure to estimate the shared lane permitted left-turn parameters. In addition, the uniform delay of the permitted left turn is
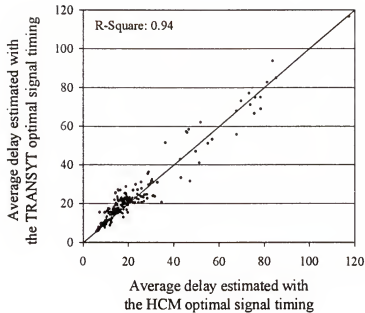
Figure 5-9: Delay comparison between TRANSYT-7F and the proposed method when permitted movement is involved

estimated based on the identical modeling approach to the one used in the protected-phase cases. The estimation of the shared lane permitted left-turn parameters should consider complicated drivers' behavior, such as lane changing of through vehicles based on drivers' preferences. The proper estimation of the lane use of through vehicles plays an important role. It should be noted that the HCM uniform delay equation has the problem with the permitted left turns. New methods that improve the estimation of the permitted left-turn parameters and uniform delay of permitted left turns are required for better design of signal timings.

5.3.2 Control Delay Equalization

A set of the signal timings designed based on the critical lane group control delay equalization strategy was obtained. In order to demonstrate the equalized control delays

of all critical lane groups, the highest control delays among the critical lane groups on major and minor streets were plotted and are presented in Figure 5-10.
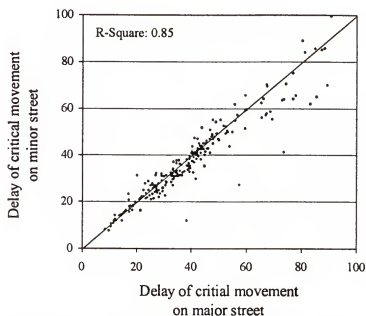
Figure 5-10: Delay comparison between the critical lane groups on major and minor streets

When traffic volume on a major street is high, the control delay of the critical movement on a minor street becomes lower than the one on the major street (a plot appears on the right-hand side of the one-to-one line). It is because of the difference in the marginal delay-changing rates (unit delay change per unit green time change). When the marginal delay-decreasing rate of the critical lane group on a major street is higher than the marginal increasing rate of the critical lane group on a minor street, the strategy introduces slightly higher delay to the major street. The R-square value between critical lane groups on major and minor streets is 0.8524. The comparison was made only between the HCM delays of the critical lane groups of those streets. Comparison

between the other models cannot be made since no existing model employing the control delay equalization strategy exists.

5.4 Maximum Green Time Optimization

The maximum green time parameters designed in the previous chapter by the proposed method, the combination of the HGA computational procedure and the modified NCHRP method, were compared to the ones designed by the other methods. The hypothetical intersection and volume conditions used in the maximum green time parameter design procedure, presented in Figure 4-12, were employed in the comparison test. Lin's method and Kell and Fullerton's method were selected to compare with the proposed method. A set of CORSIM runs were made to plot the three-dimensional charts comparable to Figures 4-14 through 4-21, which were obtained based on the proposed method, the HCM-based design method. The three-dimensional charts were plotted based on the CORSIM simulation runs for the four different volume levels: 675, 780, 820 and 900 vphpl (vehicle-per-hour-per-lane). The three- and two-dimensional charts illustrating the CORSIM delay-changing trend over different maximum green time parameters at those four volume levels are presented in Figures 5-11 through 5-18.

Those charts illustrate the CORSIM optimal maximum green time parameters that minimize the overall average delay, appeared as a pseudo concave-shaped surface. The optimal maximum green time parameters that minimize the average delay of CORSIM are 20, 30, 35 and 95 sec for volume levels 675, 780, 820 and 900, respectively. Based on Lin's and Kell and Fullerton's methods, another set of the maximum green time parameters was designed. Their methods utilize the pretimed optimal signal timings.
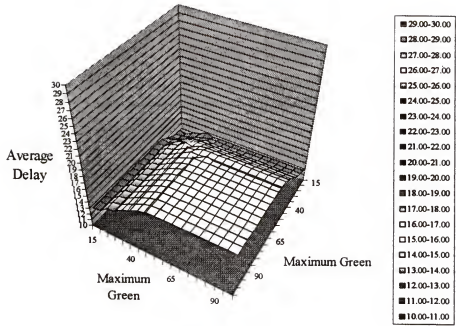
Figure 5-11: Delay changing trend (3D) over maximum green times based on the CORSIM with 675 vphpl
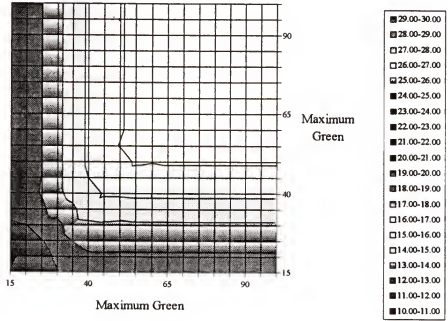
Figure 5-12: Delay changing trend (2D) over maximum green times based on the CORSIM with 675 vphpl
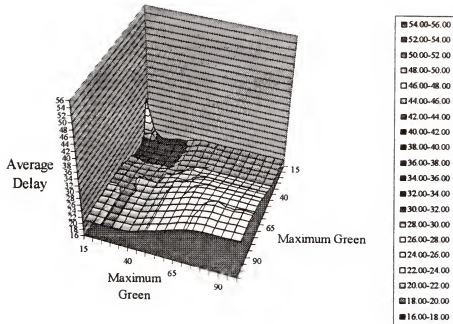
Figure 5-13: Delay changing trend (3D) over maximum green times based on the CORSIM with 780 vphpl
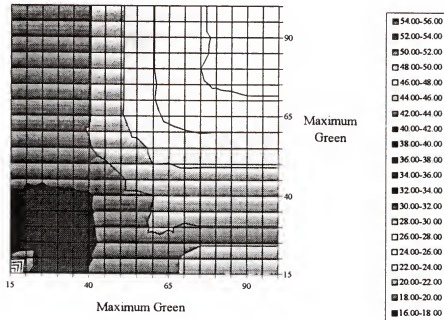
Figure 5-14: Delay changing trend (2D) over maximum green times based on the CORSIM with 780 vphpl
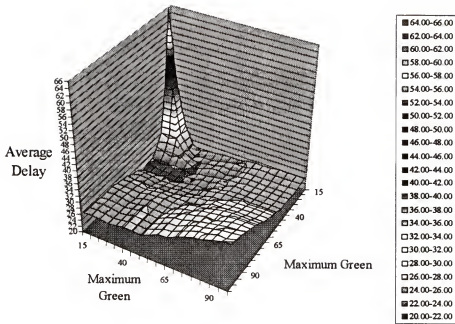
Figure 5-15: Delay changing trend (3D) over maximum green times based on the CORSIM with 820 vphpl
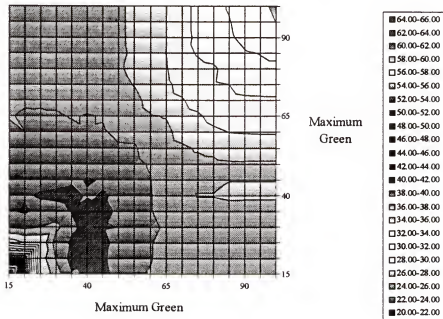
Figure 5-16: Delay changing trend (2D) over maximum green times based on the CORSIM with 820 vphpl
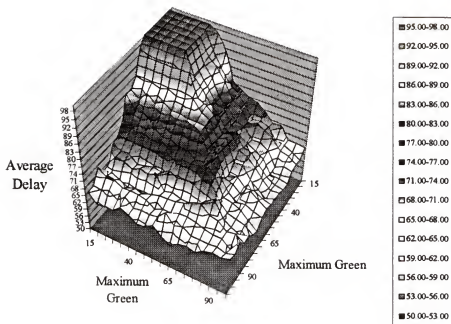
Figure 5-17: Delay changing trend (3D) over maximum green times based on the
CORSIM with 900 vphpl
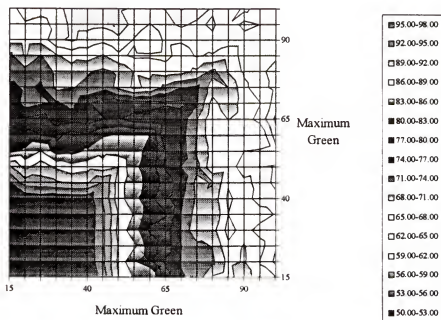
Figure 5-18: Delay changing trend (2D) over maximum green times based on the
CORSIM with 900 vphpl

Detailed description on those methods is provided in Chapter 2. The optimal values of the maximum green time parameters designed by the different methods were compared to the CORSIM optimal. The results are summarized in Table 5-4.

Table 5-11: Comparison of the maximum green times designed by different methods

Methods	Approach volume			
	675	780	820	900
CORSIM	20	30	35	95
Proposed Method	20	30	35	45
Lin	30	38	40	56
Kell and Fullerton	28	39	41	63

For the design of the maximum green parameters based on Lin's and Kell and Fullerton's methods, the optimal set of the green times for pretimed operation was calculated by SOAP2K with the delay minimization strategy. The optimal pretimed green times at the four-selected different volume levels are 20, 28, 30 and 46 seconds, respectively.

The results show that the proposed method, a combination of the modified NCHRP method and the HCM-based evaluation, provides the best performance of design. The difference of the maximum green times at 900 vphpl between CORSIM and the proposed method is due to the performance of the HCM delay model for oversaturated conditions. The CORSIM simulation results showed that, when the v/c ratio becomes larger than one, the maximum green time parameter should be increased steeply. This is the oversaturated condition. No better analytical delay estimation model

than the simulation model exits. In addition, in practice, pretimed operation is usually implemented when the operational condition of an intersection is nearly oversaturated.

CHAPTER 6
TIME-STEP-BASED ESTIMATION FOR PERMITTED LEFT TURNS

Although the HCM provides easy-to-follow manual computational procedures, it does not ensure the most accurate results because of the simplifying assumptions that must be made to adapt the procedures to the HCM worksheets. This is especially true of the procedures for analyzing permitted left turns, where it has been demonstrated that more detailed models can achieve more reliable results than the HCM [44]. For example, it was shown in Chapter 5 that the proposed optimization scheme gave superior results to TRANSYT-7F overall. However, because of its more detailed treatment, TRANSYT-7F performed better in some cases with phases involving permitted left turns because the proposed scheme was constrained to the HCM procedure (see Figures 5-9)

A time-step-based method was developed as a part of this study to overcome the deficiencies in the HCM model and to meet the need for proper estimation for permitted left-turn parameters. This chapter presents the development of the time-step-based method and the tests performed using the method.

The concept of time-step analysis is very simple. The signal cycle is divided into a series of one-second intervals and a separate analysis is performed on each interval. The number of vehicles waiting to be serviced on each approach at the end of each second is determined by adding the number of vehicles that arrived during the second to the number waiting at the beginning of the second, and subtracting the number that departed during the second. The advantage of time-step analysis is that it is able to

recognize changes in the arrival and departure parameters that occur from second to second. The disadvantage is that the whole analysis must be repeated for every second of the cycle, thereby losing the "easy to follow" characteristics that have traditionally been required of HCM procedures.

As the HCM has evolved, it has gradually moved away from oversimplified procedures into those that are demonstrably more accurate. The time-step concept is explored here to lay the groundwork for an important enhancement to future versions of the HCM. Specific topics covered in this chapter include (1) shortcomings of the HCM model and justification for an alternative approach, (2) the structure of a time-step procedure, (3) a detailed performance estimation model based on a time-step procedure and (4) a summary of tests carried out to assess the effectiveness of the time-step procedure.

6.1 Problem of the Permitted Left Turn in the HCM Methodology

The HCM suggests to adjust the saturation flow rate of a lane group containing a shared permitted left-turn movement by multiplying a left-turn adjustment factor to that whole lane group and applying the adjusted saturation flow rate to a whole green period [1, 7]. A permitted left-turn queue accumulated polygon drawn for the exclusive left-turn case based on the HCM is presented in Figure 6-1(a). Since the HCM uniform delay is obtained with the queue-accumulated diagram with the uniform arrival, it is identical to the area of the queue-accumulated polygon (QAP). The QAP drawn based on the time-step based method scheme for the exclusive left-turn case is provided in Figure 6-1(b). In Figure 6-1(c), the QAP drawn for a single shared left-turn lane case is presented.
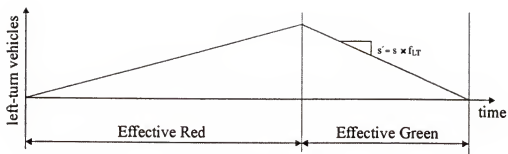
(a) The HCM method
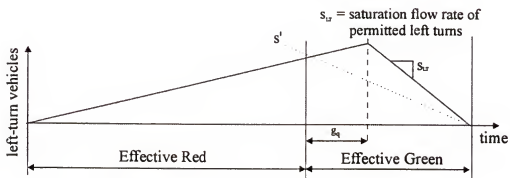
(b) Time-step-based method for exclusive lane case
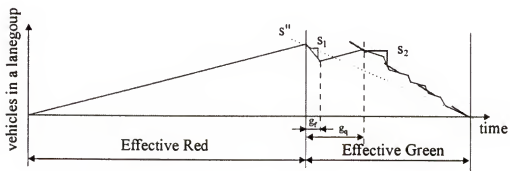
(c) Time-step-based method for shared lane case

Figure 6-1: Queue accumulated polygons

The uniform delays estimated based on the HCM and the time-step-based methods are the areas of triangles shown in Figure 6-1(a) and 6-1(b), respectively. The HCM adjusted saturation flow rate is presented in Figure 6-1(b) with s' to visualize the difference in uniform delay estimated by those two different methods. When the opposing movement uses green time, permitted left turns should experience delay. However, the HCM method failed to measure such portion of delay. The upper area of the triangle above the s' line illustrates the amount of uniform delay underestimated by the HCM method.

For the shared permitted left-turn case, it is difficult to visualize the HCM tendency of underestimation through QAP due to the complexity existing with g_f and g_q. In Figure 6-1(c), two different triangles could be found when QAP was divided by the HCM-adjusted saturation flow rate, s'. The lower and upper ones represent delay overestimated and underestimated by the HCM method, respectively. It should be noted that the HCM-adjusted saturation flow rate, s', makes the upper triangle be larger than the lower one since the slope s' becomes steeper when g_f becomes large. When g_f is identical to g_q, the QAP changes its shape to the one presented in Figure 6-1(b).

The QAP is developed for the uniform delay estimation under the assumptions of uniform arrival and the steady state condition. In the steady state condition, traffic volume should always be less than or equal to the capacity. It should be noted that when the HCM-adjusted saturation flow rate becomes lower, which yields the flatter slope of s', the HCM uniform delay should become smaller. This is because the height of the triangle becomes low in order to satisfy the steady state condition.

6.2 Development of the Time-Step-Based Method

The time-step-based method was developed to overcome the deficiency of the HCM permitted left-turn performance estimation. The proposed method was developed primarily for better estimation of average lengths of traffic-actuated phases associated with permitted left turns. The time-step based method is the one tracing permitted left turns and their opposing movements at each time step. It is a mezoscopic simulation approach utilizing the analytical method.

The method uses the queue-accumulated polygon drawn based on time-step-based activities as shown in Figure 6-1(b) and 6-1(c). Since the average length of traffic-actuated phases is affected by permitted left turns on a shared lane, it was interested to determine P_L in the time-step-based method developed in this research. Two different equilibrium assumptions were used in the determination of P_L. They are as follows:

1. Through vehicles equalize their v/c ratios on exclusive and shared lanes if the v/c ratio of through movement on exclusive lanes is higher than the one on a shared lane.

2. Through vehicles equalize the control delays of an exclusive lane and shared lane if the delay of through movement on the exclusive lane is higher than the one on a shared lane.

The P_L estimation based on the strategies of v/c and delay equilibrium between vehicles on a shared lane and exclusive lanes has been researched [44]. However, it failed to focus on the performance of through vehicles, which determines the P_L value. The time-step-based method determines P_L based on those equilibrium conditions specified above. This time-step-based method was implemented at the inside of the NCHRP iterative structure, which is designed for the convergence of phase lengths of an

actuated controller. The time-step-based computational procedure cooperated into the NCHRP computational structure is briefly illustrated in Figure 6-2. The following subsections describe the models developed for the time-step-based method.

6.2.1 P_L Estimation

The estimation of proper P_L values plays an important role in the estimation of performance of the shared permitted left-turn lane. The original and modified NCHRP methods employ the 1985 HCM models to determine the P_L value [18]. The next version of the HCM published in 1994 was available when the NCHRP 3-48 project was in progress. However, the old version of HCM models was incorporated into the NCHRP method in order to avoid the oscillation in convergence. The 1994 HCM model introduces the oscillation in the phase length convergence with the NCHRP computational structure. This is because the 1994 HCM suggests computing the permitted left-turn parameters by considering those as the functions of green times, but green times are changed when those parameters were changed within the NCHRP computational structure throughout the iterations. The problem remains with the recent versions of the HCM, published in 1997 and 2000. They utilize the same approach used in the 1994 HCM for the permitted left-turn-parameter estimation. This yields the 1985 HCM model to be the most suitable HCM model in terms of the cooperation with the NCHRP method. However, the model estimates the parameters in an oversimplified way and is too old.

The time-step-based method estimates the P_L value with the iterative manner by searching for the through vehicle proportion that the performance of through vehicles on the shared lane becomes equal to or slightly higher than the one on exclusive lanes. It estimates the P_L value based on the assumption that the performances of through vehicles
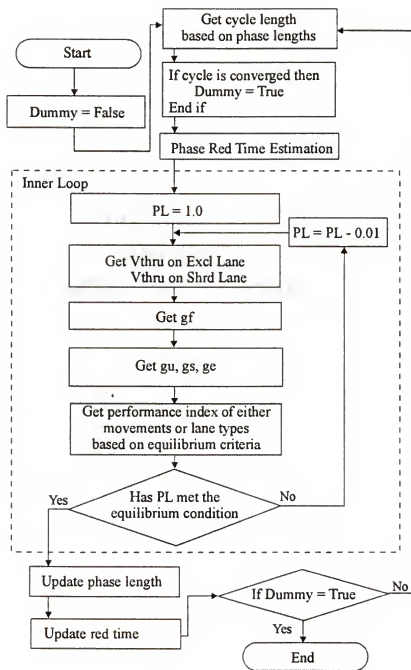
Figure 6-2: Flow chart for the time-step-based method

on those lanes tend to be equalized. Previous studies have adopted the approach that through vehicles using one of two different types of lanes, shared and exclusive lanes, by making their performance on these lanes be nearly close [44]. The v/c ratio and delay of the through vehicles are selected as the performance index in this study to determine such equilibrium conditions. Proper estimation of P_L requires the estimation of (1) performance indexes (either delays of v/c ratios) of through movements on two different types of lanes and (2) permitted left-turn parameters for the estimation of those performance indexes. The permitted left-turn parameters include (1) green time available to through vehicles prior to the first left-turn vehicle arriving, g_f, (2) the queued time taken to clear the queue of opposing movement, g_q, and (3) the unblocked time, g_u, available to left-turn vehicles after g_q.

In the time-step-based method, the P_L value is initially set to one, implying defacto left-turn lane of a shared lane. This yields the high value of performance index, such as v/c ratio or control delay, of vehicles on exclusive lanes, and the zero on a shared lane. By decreasing the P_L value by 0.01 per each analysis step, one percent through vehicles is assigned to a shared lane. The final P_L value is determined by (1) continuing to decrease the P_L value and (2) analyzing the performance index of those two different lane types until the performance index of exclusive lanes becomes equal to or lower than the one of a shared lane. This iterative procedure is illustrated in the flow chart presented in Figure 6-2.

The performance equilibrium procedure requires the higher level of accuracy at the vehicle performance estimation. Since no analytical models that estimate the performance indexes such as v/c ratio and control delay of through vehicles on a shared

lane currently exist, a set of models that estimate those performance indexes was developed to assist the time-step-based method. The delay estimation model developed to assist the time-step-based method overcomes the problem found in the HCM model as well.

6.2.2 g_f Estimation

Free green, g_f, is the time that the first left-turn vehicle arrives after the start-up lost time occurring after the start of green time. In the time-step-based method, g_f is estimated based on the geometric distribution function. When X is a random variable representing the number of failures (through vehicles) before a success (the first arrival of a left-turn vehicle on the stop line) occurs, it follows a geometric distribution function [45]. Thus, the expected value of X can be estimated by the following equation:

$$E[X] \;=\; q/p \;=\; [1 - P_L]/P_L \tag{6-1}$$

where

p = probability of left-turn vehicle arrival on a shared lane and

q = probability of through vehicle arrival on a shared lane.

g_f values estimated based on the P_L values are presented in Figure 6-3. Traffic engineers have accepted this geometric distribution-based probabilistic approach for the estimation of RTOR [5].

6.2.3 Left-Turn Saturation Flow Rate Estimation

Saturation flow rate of permitted left turns on a shared left-turn lane was estimated through the model developed by Jer-Wei Wu at the University of Florida in 1996 [46]. He developed the regression-based models for permitted left-turn movement
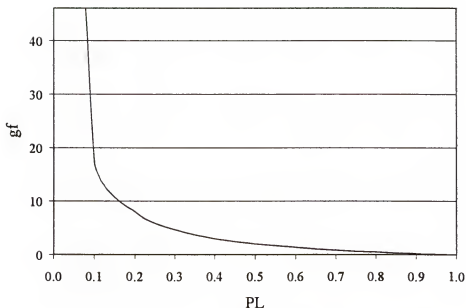
Figure 6-3: Free green estimated based on the percent left turns on a shared lane

on a shared lane for the period of g_u with the NETSIM simulation data. This model overcomes the problem of zero saturation flow rate experienced with over 1400 vph opposing traffic by the 1985 and 1994 HCM models. The permitted left-turn saturation flow rate should be zero when the opposing flow rate is nearly saturated. The model used in the time-step-based method is provided below:

$$S_s^{LT} = 3600/t_f - (0.2t_c + 1.12)v_o + (2t_c + 171)10^{-5} v_o^2 - (1007 - 66t_c)10^{-9} v_o^3 \quad (6\text{-}2)$$

where

t_f = average discharging headway (sec/veh) and

v_o = opposing vehicles.

In the time-step-based method, 2.1 seconds were used as the average discharging headway of permitted left-turn vehicles during unsaturated green. It is the default value

used in the HCM for the exclusive lane and protected phase cases. It was used to reflect the unsaturated green time situation and to make it independent from the effect from the opposing traffic. It reflects the average left-turn headway observed on shared lanes when there is no opposing traffic.

6.3 Development of the Performance Estimation Models

Two different types of equilibrium conditions were developed for the estimation of P_L. They include (1) the equilibrium of v/c ratios of through vehicles on exclusive and shared lanes and (2) the equilibrium of control delay of through vehicles on exclusive and shared lanes. The estimations of the v/c ratio and control delay of through vehicles on a shared lane were made for the estimation of P_L.

6.3.1 v/c Ratio Estimation Models

For the estimation of v/c ratios of through vehicles on two different types of lanes, shared and exclusive, the capacities of those lanes for through movements were explicitly estimated. For the exclusive lane case, the capacity can be estimated easily with the HCM procedure. The HCM estimates the capacity of through movement by the following equation.

$$c_e = s_e \big(g/C \big) \tag{6-3}$$

where

c_e = capacity of through movement on an exclusive lane,

s_e = saturation flow rate of through movement,

g = effective green time and

C = cycle length.

Capacity estimation for through movement on a shared lane introduces complexity due to the interaction between left-turn and through vehicles on that lane. Left-turn vehicles block the lane at the stop line while they are waiting for gaps in their opposing flow. The time portions that through vehicles can flow should only be considered only in the capacity estimation of through vehicles. The equation used in the time-step-based method is

$$c_s = s_s^{TH}\left(g_f + g_u - g_L\right)\big/C \qquad (6\text{-}4)$$

where

c_s =   capacity of through movement on a shared lane and

g_L =   time portion used by left-turn vehicles during g_u.

g_L is the average time spent by permitted left turns during the unopposed green. It was estimated by the following equation in the time-step-based method:

$$g_L = w\,LTC, \quad \left(0 \le g_L \le g_u\right) \qquad (6\text{-}5)$$

where

w =   average gap waiting time of a left-turn vehicle at the stop line and

LTC =   the expected number of left-turn vehicles arriving during a cycle.

The value of w was estimated based on the statistical approach, while LTC was obtained by dividing the hourly left-turn traffic volume by a cycle length as computed in the HCM procedure. Let X be a random variable, the number of gaps that a left-turn vehicle needs to wait before it finds a bigger gap than the critical gap. In the Bernoulli trial, let success and failure in trial be the one if the gap is bigger and smaller than the critical gap, respectively [45]. The interest is given to the number of failures expected to be experienced prior to the first success. In other words, the average waiting time of left-

turn vehicles at the stop line is the function of the number of gaps it faces before it finds a bigger gap than the critical gap. Thus, w should be estimated based on the geometric distribution function, which the random variable X follows.

The events of success and failure of the Bernoulli trial occur by following the exponential distribution function, since they are headways of opposing traffic. Let p and q be the probabilities of success (a gap in opposing flow is bigger than or equal to a critical gap) and failure (the gap is smaller than a critical gap), respectively. The p and q were estimated through the cumulated density function of the exponential distribution.

$$p = \exp(-t_c/T) \tag{6-6}$$

$$q = 1 - \exp(-t_c/T) \tag{6-7}$$

where

T = average headways of opposing traffic and

t_c = critical gap.

The average number of gaps that the first left-turn vehicle should face at the stop line during g_u was estimated by the exponential distribution function and Equations 6-6 and 6-7. In time-step-based method, Equation 6-8 was used to estimate the number of gaps a left-turn vehicle should face at the stop line.

$$E[X] = q/p = [1 - \exp(-t_c/T)]/\exp(-t_c/T) \tag{6-8}$$

The average time, w, used by the left-turn vehicles at the stop line during g_u was computed by multiplying the average size of the opposing gap, T, and the average number of opposing gaps it should wait for. The w value is estimated for the time-step simulation method by the following equation:

$$w = T\,E[X] \tag{6-9}$$

Four point five seconds was used as the size of critical gap in the time-step-based method. It is the critical gap size used in the HCM for permitted left turns. With the equations developed for capacity estimation on a shared lane, the v/c ratio of through vehicles can be computed and compared to the v/c ratio of through vehicles on an exclusive lane.

6.3.2 Uniform Delay Estimation Models

The HCM control delay equation composes three different parts: uniform delay, random delay and residual queue delay [1, 7]. The first two delay portions are important in delay estimation for a given green time period, while the last part is designed for the delay affected by the previous analysis period for the successive time-series studies. The first two portions of the delay, uniform delay and random delay, were considered as the control delay, and the last term, residual delay, was neglected by assuming that there should be no residual queue left over from the previous period at the beginning of study period [7]. In the time-step-based method, uniform delay was estimated based on the newly developed time-step-based method, while random delay was estimated based on the HCM random delay equation.

Two different uniform delay models were developed for two different test purposes. The first delay model was developed to estimate the uniform delay of permitted left turns on an exclusive lane. It was used to test if the time-step-based method overcomes the deficiency found in the HCM uniform delay model. The exclusive left-turn case was considered for the purpose of comparison to the HCM estimation. This is because the HCM deals with shared permitted left turns and through movements as a lane group and computes the aggregated delay of those movements. By considering exclusive lanes only, it was expected to check the difference of those two

modeling approaches. The other uniform delay model was made for the estimation of through vehicles on a shared lane. This was done to obtain proper value of P_L and to estimate the average lengths of traffic-actuated phases operating with a shared left-turn lane. The following subsections describe the uniform delay estimation model developed for the time-step-based method.

6.3.3 Uniform Delay of Permitted Left Turns on an Exclusive Left-Turn Lane

The time-step-based uniform delay of permitted left turns on an exclusive left-turn lane is equivalent to the area of the queue accumulated polygon presented in Figure 6-4. The shape of the queue accumulated polygon changes with the permitted left-turn parameters, such as g_f, g_q and g_u, and the parameters estimated by the time-step simulation method. For example, g_q is the time that the effect from the opposing queued vehicles disappears. It is the sum of G_q, G_a and G_e. Based on the average vehicle arrival rate and the permitted left-turn saturation flow rates, the uniform delay of permitted left turns on an exclusive left turn was computed.
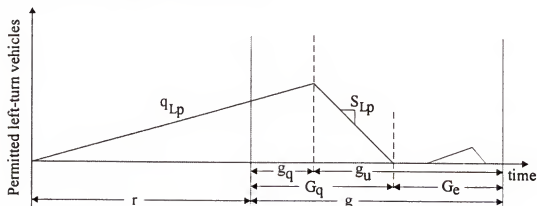
Figure 6-4: Uniform delay of permitted left turns on an exclusive lane

By definition, uniform delay is the amount of time delay experienced under the steady state condition; traffic demand volume should be lower than or equal to the

capacity in the study period. When v/c ratio of movement becomes higher than one, the random delay portion plays a much more important role, and uniform delay carries the maximum uniform delay expected when v/c ratio is equal to one. To take this into account, the maximum length of the queue that can be served during g_u was defined in the time-step-based method as the following equation:

$$Q_M^{g_u} = g_u \max\left[s_e^{LT} - q_e^{LT}, \quad 0\right] \tag{6-10}$$

where

$Q_M^{g_u}$ =   the maximum length of maximum back of queue satisfying the steady state condition,

S_e^{LT} =   saturation flow rate of left turns on an exclusive left-turn lane and

q_e^{LT} =   left-turn arrival rate on an exclusive left-turn lane.

The actual maximum back of queue was estimated through the following equation:

$$Q_M = \left(r + g_q\right) q_e^{LT} \tag{6-11}$$

where

Q_M =   maximum back of queue and

r =   effective red time.

The uniform delay was determined based on those $Q_M^{g_u}$, Q_M and other permitted left-turn parameters. The equation used in the time-step-based method is presented in Equation 6-12.

$$d_1 = \begin{cases} \left\{Q_M\left(r + G_q\right) + wG_e q_e^{LT}\right\}/\left(2q_e^{LT}C\right) & , if \ Q_M < Q_M^{g_u} \\ Q_M^{g_u}/\left(2q_e^{LT}\right) & , else \end{cases} \tag{6-12}$$

where

d_l =   uniform delay,

g_q =   saturation flow rate,

q =   average vehicle arrival rate,

C =   cycle length and

S_{LT} =   saturation flow rate of permitted left turns during unopposed green.

This uniform delay equation contains the delay that permitted left turns experienced during g_u. This portion is illustrated in Figure 6-4. The small triangle located in the period of G_e represents the uniform delay of the permitted left turns waiting for the gaps during unopposed green after G_q. This additional uniform delay, d_a, was estimated by Equation 6-13, and it was already contained in the total uniform delay estimation model presented in Equation 6-12.

$$d_a = wq_e^{LT}G_e \qquad (6\text{-}13)$$

For the estimation of random delay through the HCM equation, the capacity of permitted left turns was required for the computation of the v/c ratio. The capacity of permitted left turns on an exclusive left-turn lane was estimated by the following equation:

$$c_e^{LT} = s_e^{LT}\left(g_u/C\right) + n(3600/C) \qquad (6\text{-}14)$$

where

c_e^{LT} =   capacity of permitted left turns (vph) and

n =   left-turn sneakers per cycle.

Left-turn sneakers were estimated by the regression-based model developed by Jer-Wei Wu at the University of Florida in 1996 [46]. The model is presented in Equation 6-15.

$$n = 2.34 - 1.74 \times 10^{-3} v_o + 8.44 \times 10^{-7} v_o^2 \qquad (6\text{-}15)$$

6.3.4 Uniform Delay of Through Vehicles on a Shared Lane

The second uniform delay model was designed for the delay of through vehicles using a shared left-turn lane. This uniform delay model was used in the determination of the performance of through vehicles using a shared lane. The delay model is composed of two major portions and is presented in Equation 6-16.

$$d_1 = d_{1A} + \varphi \, d_{1B} \tag{6-16}$$

where

d_{1A} = delay of through vehicles served during

d_{1B} = delay of through vehicles served during and

φ = proportion of through vehicles served during to all vehicles served during.

The first portion of uniform delay is the one experienced by (through) vehicles served during g_f, and the second portion of the uniform delay is the one experienced by (through) vehicles served during g_u. Figure 6-5 illustrates the first portion of delay with the shaded area.
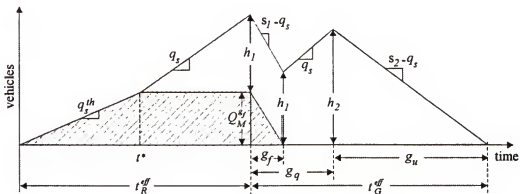
Figure 6-5: Queue accumulated polygon for a shared left-turn lane

The first term of delay equation is estimated through the following equation:

$$d_{1A} = Q_M^{g_f}\left(2r + g_f - t^*\right)\Big/\left(2q_s^{TH}C\right) \qquad (6\text{-}17)$$

where

$Q_M^{g_f}$ =   average length of queue expected to be served during g_f,

t^* =   time period that queue length is less than and

q_s^{TH} =   average arrival rate (vps) of through vehicles on a shared lane.

$Q_M^{g_f}$ is the average length of queue expected to be served during g_f, and t^* is the time that the $Q_M^{g_f\,\text{th}}$ vehicle is arriving. They were determined by the following equations:

$$Q_M^{g_f} = \min\left[rq_s^{TH}, \quad g_f\left(s_1 - q_s^{TH}\right)\right] \qquad (6\text{-}18)$$

$$t^* = Q_M^{g_f}\Big/q_s^{TH} \qquad (6\text{-}19)$$

The second term of uniform delay was estimated by the following equations:

$$d_{1B} = \left[h_1\left(r - t^* + g_f + g_q\right) + h_2\left\{g_q - g_f + h/(s_2 - q_s)\right\}\right]\Big/\left(2q_s^{TH}C\right) \qquad (6\text{-}20)$$

$$\varphi = \frac{q_s^{TH}\left(C - t^*\right)}{q_s^{TH}\left(C - t^*\right) + q_s^{LT}C} \qquad (6\text{-}21)$$

where

h_1 =   the height of the queue-accumulated polygon at the end of g_f and

h_2 =   the height of the queue-accumulated polygon at the end of g_q.

Those heights were computed by the following equations:

$$h_1 = h_2 - q_s\left(g_q - g_f\right) \qquad (6\text{-}22)$$

$$h_2 = \min\left[q_s\left(r - t^*\right) + g_q - g_f, \quad g_u\left(s_2 - q_s\right)\right] \qquad (6\text{-}23)$$

The capacity of through vehicles on a shared left-turn lane is required for the computation of random delay. It was estimated by the following equation in the time-step-based method.

$$c_s^{TH} = s_1 \left(\frac{g_f + g_u - g_L}{C} \right)$$
(6-24)

where

c_s^{TH}	=	capacity of through vehicles on a shared lane,
s_1	=	saturation flow rate of shared lane during g_f,
s_2	=	saturation flow rate of shared lane during g_u,
q_s^{LT}	=	average arrival rate of left-turn vehicles and
h_{LT}	=	average headway of left-turn vehicles.

The v/c ratio of through vehicles on a shared lane was computed based on the capacity estimated by Equation 6-24. Random delay was calculated with the estimated v/c ratio.

6.4 Test of the Time-Step-Based Method for an Exclusive Left-Turn Lane

The time-step-based method was designed in this research for better performance estimation of permitted left turns. Two tests were performed with the time-step-based method by adopting the CORSIM simulation data as surrogate for field data. The first test was to compare the control delay estimated by the time-step-based method to the one from the HCM model. It was intended to check that the time-step-based method overcomes the shortage of the HCM permitted left-turn delay estimation previously indicated. Since the HCM deals with permitted left-turn movement as a lane group with a through movement for a shared lane, an exclusive permitted left-turn lane was

considered in the test. The second test was performed with the case of a shared permitted left-turn lane operated with the traffic-actuated controller. Estimation of the average lengths of traffic-actuated phase with a shared lane was made. The obtained average phase lengths were compared to the ones estimated by the NCHRP method. During this process, the assumptions of v/c and delay equilibrium conditions were considered to split through vehicles on two different types of lanes: exclusive and shared left-turn lanes.

In the first test, the control delays of permitted left-turns from the time-step-based method and the HCM method were compared to the CORSIM simulation data. In the CORSIM simulation analysis, the sample size (the number of simulation runs) was determined based on a control scheme. First, with 10 initial runs, the average and standard deviations of such variables were found and checked if standard deviations were less than 10 percent of the averages. If they were larger than 10 percent of the averages, sample sizes of those specific cases were to be increased by adding 10 more simulation runs. However, in all cases, it turned out that 10 CORSIM runs were enough to satisfy this control scheme. Different simulation runs use different sets of random seed numbers predefined.

It is critical to force the simulation conditions to be as close as possible to the assumptions those analytical models possess in order to prevent effects of the CORSIM simulation model. The same condition used in Chapter 6 was adopted in the test. This includes 4.5 seconds of critical gaps in record type 145, no left-turn jumper in record type 140, and two left-turn sneakers in record type 141. Permitted left turns on an exclusive left-turn lane were considered to demonstrate the difference between estimations made by the HCM and the time-step-based method. This is because the HCM deals with

permitted left-turn movement with a lane group containing through movement. The HCM provides aggregated control delay of the lane group but not one for permitted left turns only. The case of permitted left turns opposed by a shared left-turn lane was excluded in the test due to its complexity in computation.

6.4.1 Test Procedure

A hypothetical intersection was designed to test the performance of the proposed delay estimation model. The intersection is shown in Figure 6-6. Northbound and eastbound approaches have permitted left turns, while through traffic is set only on southbound and westbound approaches. Pretimed signal operation was applied since the purpose of the test is to check the difference in the two different delay estimation approaches. The basic condition was determined as shown in Figure 6-6.

30 mph

480

40 mph   78   520
          702   40 mph

72 648

30 mph

Phase 1 | Phase 2

Suburban area
Saturation flow rate = 1800 vph
% Truck = 0
PHF = 1.0
Lost Time = 3 seconds
Minimum green = 10
Maximum green = 46
Unit extension = 3
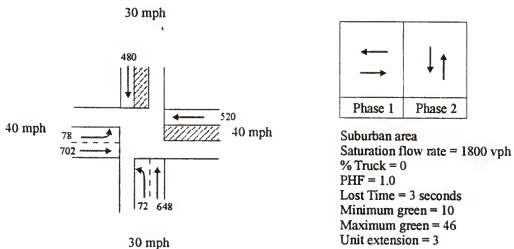
Figure 6-6: The hypothetical intersection used in the test of the time-step-based method

Based on the standard condition presented in Figure 6-6, various traffic conditions were generated by multiplying five volume-increment factors, 0.7, 0.8, 0.9, 1.1, 1.2 and 1.3, by changing the directional balances to 50:50, 55:45, 65:35 and 70:30 and the left-

turn percent to 0.05, 0.15, 0.20 and 0.25. Various control conditions were considered in tests with a cycle length of 60, 90, 120 and 150 sec. A total of 120 different cases of traffic conditions of permitted left-turns was used in the test.

6.4.2 Comparison of the Control Delays Estimated

The control delays estimated by the proposed method and the HCM method are illustrated in Figure 6-7 and Figure 6-8. The HCM control delay compared to the one from the simulations is shown in Figure 6-7. They show that the HCM delays underestimate the delay of permitted left turns as discussed. The control delays estimated by the time-step-based models are close to or higher than the one-to-one relation to the simulated control delay. Those charts visualize the power of the time-step-based method in delay estimation for permitted left-turn movement. The R-square values of the HCM and the time-step-based methods to the simulated control delay were 0.47 and 0.77, respectively.
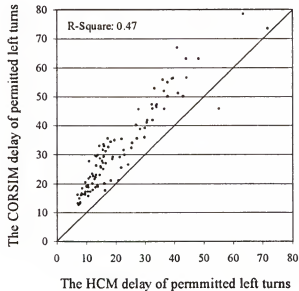
Figure 6-7: Delay comparison between CORSIM and the HCM method
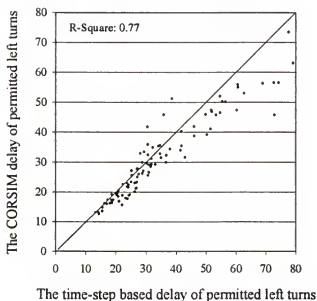
Figure 6-8: Delay comparison between CORSIM and the time-step-based method

It shows that the time-step-based models follow the similar trend of permitted left-turn delay that the microscopic simulation analysis provides. It should be noted that the CORSIM program provides optimistic performance of permitted left turns [46]. It would be expected that the delay of permitted left turns in the field should be at least equal to or a little higher than the one from CORSIM. It concludes that the proposed method performs better than the traditional HCM method since the simulation method is the best way to estimate the control delay of permitted left turns due to the complexity.

6.5 Test of Time-Step-Based Method for a Shared Permitted Left-Turn Lane

Average lengths of phases associated with permitted left turns were estimated by the NCHRP method. With a hypothetical intersection data set, the time-step-based method and the HCM method of permitted left-turn performance estimation were separately applied in the NCHRP phase length estimation procedure. The phase lengths

estimated by those procedures were compared to the phase length estimated by the microscopic simulation analysis. The CORSIM simulation program was used as a tool of simulation analysis. Multiple CORSIM runs were made, and the average phase lengths were obtained through TSDREADER.

6.5.1 Test Procedure

A hypothetical intersection used for the test of the time-step-based method is provided in Figure 6-9. Northbound and eastbound approaches have a shared left-turn lane, and their opposing approaches, southbound and westbound ones, have a single through lane, respectively. The various traffic conditions were considered in the test, and the conditions were generated with the rules used in the test of protected phases. Two left-turn sneakers per cycle was considered.
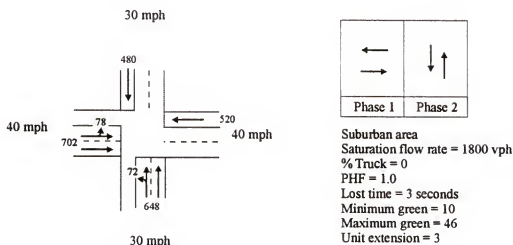
Figure 6-9: The hypothetical intersection used in the test of the time-step-based method for the shared lane case

Two different types of equilibriums were tested. They include the following:

1. Delay equalization between through vehicles on shared and exclusive lanes. Figure 12 and Figure 13 show phase length and P_L values estimated by the method.

2. v/c ratio equalization between through vehicles on shared and exclusive lanes. Figure 14 and Figure 15 show phase length and P_L values estimated by this method.

When delay equilibrium is used, different trends of P_L is observed; P_L reaches to 1.00 more rapidly than one observed with the v/c ratio equilibrium because of the left-turn sneakers. The left-turn sneakers increase the capacity of the shared lane and introduce more through vehicles into the shared lane under the v/c ratio equilibrium strategy. On the other hand, they increase delay on a shared lane, so less through vehicles use the lane. In this test study, two left-turn sneakers per cycle were applied.

6.5.2 Comparison of the Phase Lengths Estimated

The condition of the v/c ratio equilibrium of through vehicles on a shared left-turn lane and exclusive through lanes was considered in the determination of P_L. Based on the volume distribution estimated on that criterion, average lengths of traffic-actuated phases were estimated through the NCHRP method. Figure 6-10, Figure 6-11 and Figure 6-12 were drawn to visualize the difference in phase length estimated by different models. Figure 6-10 shows the phase length comparison between the CORSIM and the NCHRP methods with the 1985 HCM. Figure 6-11 and Figure 6-12 show the phase length comparison between the CORSIM and time-step-based methods with v/c equilibrium and the control delay equilibrium strategies, respectively.

Figure 6-10: Phase-length comparison between CORSIM and the NCHRP method

Figure 6-11: Phase-length comparison between CORSIM and the proposed v/c equilibrium method

Figure 6-12: Phase-length comparison between CORSIM and the proposed delay equilibrium method

Figure 6-11 illustrates that the 1985 HCM method underestimates the average lengths of phases. Phase lengths estimated based on the 1985 HCM method are less than the ones obtained from CORSIM. Figure 6-12 shows that the v/c ratio equilibrium models introduce more one-to-one relationships, but it is a little biased to the right side. The v/c equilibrium time-step-based method yields a little longer phase length than the CORSIM. Figure 6-12 shows the best one-to-one relationship among three models tested with CORSIM with a little less R^2 value, 0.8346, than the one of the v/c ratio equilibrium model. The v/c ratio equilibrium model yields the best R^2 value, measure of one-to-one relationship, 0.8457. Through the comparison test, it has been shown that the equilibrium models provide better estimation of the average phase length. Those two models provide a better linear relationship to the CORSIM than the 1985 HCM model, whose R^2 is 0.5271. The v/c ratio equilibrium model provides a better linear relationship with

CORSIM, and the delay equilibrium model provides the best one-to-one relationship with the CORSIM results.

6.5.3 Comparison of the P_L Estimated

The P_L values estimated through those models were compared to the one from the CORSIM. The 1985 HCM provides the most one-to-one linear relationship, while both equilibrium models with time-step-based method produced different relationships than that. Figure 6-13 shows the relationship between the 1985 HCM results and the CORSIM results. Figure 6-14 and Figure 6-15 present the relationship between the v/c equilibrium results and the CORSIM results and between delay equilibrium results and the CORSIM results, respectively.

All those models show that the P_L value converges to one more rapidly than the simulation model does. This apparently has been shown with the two equilibrium models. This is because of the optimistic behavior of CORSIM with the permissive left-
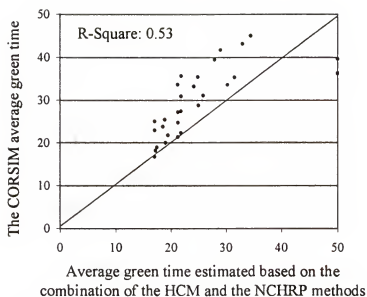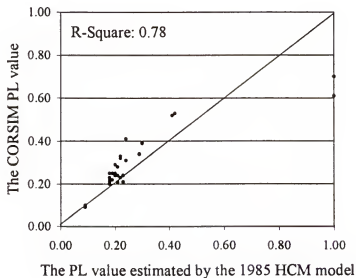

Figure 6-13: P_L comparison between CORSIM and the NCHRP method
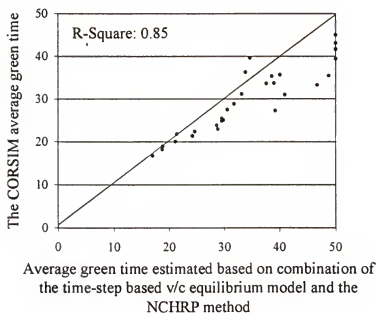
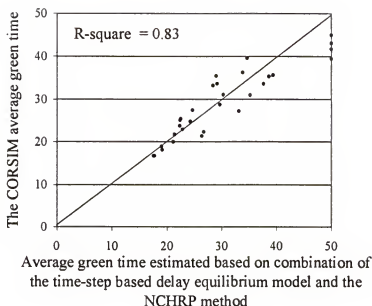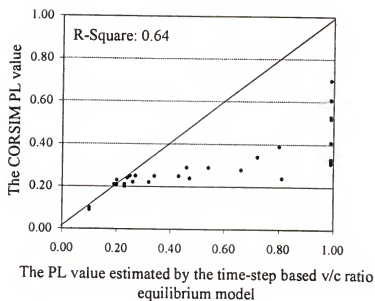Figure 6-14: P_L comparison between CORSIM and the proposed v/c equilibrium method
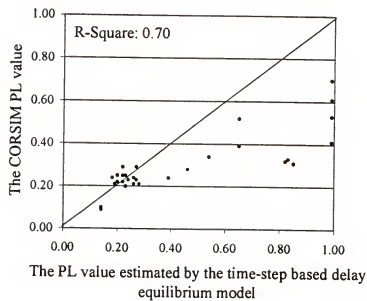
Figure 6-15: P_L comparison between CORSIM and the proposed delay equilibrium method

turn vehicles. Although 4.5 seconds of critical gap were constantly applied, CORSIM considered the left-turn drivers become aggressive, and the drivers responded to the shorter gaps than their critical gaps assigned when their waiting time becomes longer.

Also, the follow-up headways of permitted left turns were not considered in the models, while the simulation model was considered. The P_L value it yields becomes smaller when P_L is larger than 0.3. With the 1985 HCM method, the P value is higher than the one from CORSIM when it ranges between 0.2 and 0.6, while other equilibrium models produce lower values of P_L, compared to it.

The time-step-based model shows the expected trend of permitted left-turn delay, while the HCM underestimates the delay. A statistical test is recommended with a larger data set. In terms of phase length estimation, the time-step-based method with a delay equilibrium model performed better than the other, while its P_L shows a different trend than the one from the CORSIM runs. It is recommended to calibrate a new model that estimates a new P_L value based on delay equilibrium P_L and v/c equilibrium P_L.

CHAPTER 7
CONCLUSIONS AND RECOMMENDATIONS

This dissertation has presented (1) the computational procedures that design the pretimed signal timing and actuated maximum green time parameters for isolated intersections based on the HCM performance evaluation method and (2) the comparison made among the delays introduced with the signal timing parameters designed by the proposed computational procedure and the existing signal timing design models. The comparison has demonstrated that the signal timings designed by the existing computer-based models used in practice are different from the HCM optimal. This suggests that the improvement expected by the HCM has not fully satisfied the existing practical engineering method.

In addition, the dissertation presented the improvements made in the estimation of (1) the average lengths of actuated phases, (2) the permitted left-turn parameters and (3) permitted left-turn delay. In comparison with simulation results, it has been shown that the proposed computational procedure, the HCM-based design method, is a better tool to design signal-timing parameters than existing models. The conclusions and recommendations addressed in this chapter present the results of the research.

7.1 Conclusions

Four major conclusions of the research and their support considerations are presented below.

First, it has been found that the proposed computational procedure is able to find the HCM optimal signal timings. This is supported by the following:

1.   For pretimed operation, it has been demonstrated that the signal-timing parameters designed by the proposed computational procedure, the combination of the HGA searching and the HCM performance evaluation procedure, achieves the optimization of the control parameters, cycle and split.

2.   For traffic-actuated operation, it also has been demonstrated that the proposed computational procedure, the combination of the HGA searching procedure, the modified NCHRP3-48 method and HCS, is able to find the HCM optimal maximum green time parameters.

Second, in comparison with simulation results, it has been demonstrated that the signal timings designed by the existing models used in practice are not the HCM optimal and that better signal timing can be found through the proposed computational procedure. This demonstrates that it is possible to improve the practice of signal timing design using the proposed procedures. The following considerations support this conclusion:

1.   It has been demonstrated that the pretimed signal timings designed based on the proposed computational procedure introduce lower delays than those from the existing models.

2.   It has been demonstrated that the actuated maximum green time parameters designed based on the proposed method, the combination of NCHRP3-48 method and the HCM evaluation procedure, introduce lower delays than those from the existing models.

Third, the NCHRP method may be improved with the modification in the definition of the queue service time and the volume adjustment for shared right turns. The following considerations support this conclusion:

1. In comparison with the phase lengths estimated by the original and modified NCHRP methods, it was shown that the modified method is based on a more sound analytical approach than the one used in the original method. For example, the original method defines the queue service time as the length of the QAP bottom, which indicates the time required to clear the queue, and considers all RTOR vehicles as a through lane group. The modified method defines the queue service time as the green time affected by queued vehicles and excludes the RTOR vehicles from the lane group volume.

2. In comparison with simulation results, the modified method provides better estimation of the average phase lengths than the original method.

Fourth, the current HCM delay estimation procedure underestimates the uniform delay of the permitted left-turn movements, and the time-step simulation method provides improved estimation of the permitted left-turn parameters and delay. The following considerations support this argument:

1. It has been demonstrated that the current HCM procedure underestimates the uniform delay of permitted left turns due to the failure in the explanation of the delay experienced by the permitted left turns waiting to stop at the stop line until the opposing queue is cleared during the green phase.

2.   In comparison with simulation results, the proposed time-step based method has provided better estimation of delay than the HCM method.

3.   With the combination of the updated NCHRP method, it has been shown that the time-step-based method provides better estimation of the average phase length than the HCM for the average length estimation for the phases involved with permitted left turns.

7.2 Recommendations

Two sets of recommendations have been presented in this section for the improvement in the signal timing design practice and for further research:

First, it is recommended that the existing computer-based signal timing design models should be based on the HCM saturation-flow adjustment procedure. The existing models use the simplified saturation flow rate values specified by users and consistently apply those in the design procedure without adjustment based on the changed signal timings. The HCM has been developed over many decades and is accepted as the standard by many states including the state of Florida. It is recommended that the saturation flow rate estimation of the models be computed based on the HCM procedure. This will enhance the credibility of the signal timing design procedures performed in the traffic engineering profession in many states in the nation.

In addition, it is recommended that a new version of HCS be developed for the purpose of efficient research. In the comparison test, the HCS runs were made one by one because the program does not support either multiple runs or being invoked from other programs. Since many researchers compare their own models to the HCM, a version of the HCS that could be called and executed by other programs would significantly contribute to their research.

It is recommended that the HCM procedure should adopt the iterative computational procedure for the estimation of permitted left-turn parameters. While the HCM has traditionally employed single pass worksheets, the precedent for iterative procedures has already been set with the introduction of the NCHRP method for the estimation of the average actuated phase lengths. The iterative computation is highly recommended for the proper estimation of P_L. The HCM worksheet procedure has limitations that preclude the proper treatment of the complex nature of permitted left turns.

It is recommended for future research that optimization of the phase sequence by the HCM-based method be studied. Since the proposed HGA searching method provides wide flexibility in the shape of the objective function, various strategies and control parameters including phase sequence based on the HCM evaluation can be tested. Since the phase sequence is one of the important parameters affecting intersection control, the identification of the HCM optimal phase sequence would be meaningful to the traffic engineering profession. The proposed HGA procedure possesses the functionality of searching for both signal timings and phase sequence based on the HCM evaluation procedure without major modification in its computational structure.

APPENDIX

TRAFFIC MODEL MARKUP LANGUAGE (TMML)

TMML is the Extensible Markup Language (XML) based traffic model database developed during the study [47, 48]. XML is a logical extension of HTML (Hyper Text Markup Language), which has been used for the Internet-based data transferring, with its own context-specific vocabulary [49, 50, 51, 52]. XML has been accepted as means of transferring data between two systems or users who deal with the same data but in different formats [47, 48]. The "X" in XML denotes "extensible," which differs from "extended" in the sense that you have to provide your own extensions. For example, a tag called <Volume> would be considered foreign to an HTML document. It has been useful by establishing its significance. A set of such tags used in TMML and their meaning are defined in the following section.

Specifically, the Traffic Model Markup Language (TMML) has been developed to facilitate sharing of data between traffic modeling software products. TMML is a fully XML-compatible markup language prescribing the class structure and data element tag names required to represent traffic model data in a "self-describing" format. The principal applications of TMML include exchanging data between traffic model software products and facilitating the compilation and presentation of results.

The TMML language is defined in terms of a collection of data structures that describe the properties of the objects associated with traffic carrying facilities. TMML

provides the structure and vocabulary to completely define a data set for various facilities and software products. It was developed at the University of Florida and has been applied extensively to software products that analyze the performance of several facilities, including freeways, urban and rural highways, and signalized arterials.

The TMML specification includes a list of recognized abbreviations intended to reduce the size of the XML tags that describe the data elements. TMML was created following the general principles of style described in IEEE Standard 1489 [53] and the TSDD, a comprehensive list of terms, which would serve all the specified simulations, developed by the ITT Systems [54].

TMML Description

An example of the TMML structure applied to a signalized intersection is presented in Figure A-4. It starts with a start tag and ends with an end tag. An attribute that defines the property of the tag should be inserted in the start tag. Between tags, all the data related to the tag can be stored. It stores traffic data in a hierarchical manner within a set of tags. For example, 'INTERSECTION' element contains 'APPROACH' elements, and 'APPROACH' element contains 'MOVEMENT' element. A sample APPROACH section is represented in Figure A-4. For the case, it starts with a start tag for an approach and ends with an end tag. An attribute that defines the directional bound of the approach should be inserted in the start tag.

Each class has several data element tags that define its properties. TMML files can be somewhat lengthy, so the principles and guidelines that govern the TMML language include

```
<?xml version="1.0"?>
<TMML Facility="Signal">
    <INTERSECTION Int= "1">
        <EWStreetName>Gator Street</EWStreetName>
        ....
        <APPROACH ID= "EB">
            <ParkingLeftYN>N</ParkingLeftYN>
            ....
            <MOVEMENT ID= "L">
                <NumberOfLns>1</NumberOfLns>
                <AdjVol>117</AdjVol>
                ....
            </MOVEMENT>
            <MOVEMENT ID= "T">
                <NumberOfLns>2</NumberOfLns>
                <UnAdjVol>546</UnAdjVol>
                <AdjVol>546</AdjVol>
                <Movts>TR</Movts>
                ....
            </MOVEMENT>
            <MOVEMENT ID= "R">
                <NumberOfLns>0</NumberOfLns>
                <UnAdjVol>117</UnAdjVol>
                <PHF>1</PHF>
                ....
            </MOVEMENT>
            <LANEGROUP ID= "Left">
                <Movts>L</Movts>
                <GCRatio>0.48</GCRatio>
                <VSRatio>0.13</VSRatio>
                <UniformDelay>31.1</UniformDelay>
                <LOS>C</LOS>
                ....
            </LANEGROUP>
            <LANEGROUP ID= "Center">
                <Movts>TR</Movts>
                <AdjVol>1326</AdjVol>
                <ArrivalType>3</ArrivalType>
                <LOS>D</LOS>
            </LANEGROUP>

            <LANEGROUP ID= "Right">
            ....
            </LANEGROUP>
        </APPROACH>
        ....
        <TIMINGPLAN Phase= "1">
            <Green>46</Green>
            <Yellow>3</Yellow>
            <AllRed>1</AllRed>
            <PHASECODES ID="EB">
                <LTCode_HCS>P</LTCode_HCS>
                ....
            </PHASECODES>
            <PHASECODES ID="WB">
                <LTCode_HCS>P</LTCode_HCS>
                ....
            </PHASECODES>
        </TIMINGPLAN>
        <TIMINGPLAN Phase= "2">
        </TIMINGPLAN>
        ....
    </INTERSECTION>
</TMML>
```

Figure A-1: Sample TMML tags

1. The root tag is always <TMML> with an attribute defining the type of facility that the data represents.

2. Each class tag has a maximum of one attribute to facilitate reading and parsing of data. That attribute distinguishes the specific instance of the class from all other instances of the same class. For example, the attribute of the APPROACH class contains EB, WB, NB and SB, which represents the directional-bound of approach.

3. The order of presentation of classes or elements within a class is not restricted. It is possible to present the classes and elements within a class in any order as long as the class structure is maintained.

4. The same class may appear more than once in a file (e.g., to separate input and output data). A given data item may appear only once in any class.

5. Class elements are represented in upper-case characters (e.g., APPROACH). Data elements within each are represented as a concise series of connected words with lower-case characters and initial capitals (e.g., MedianWidth).

6. Brevity is not an essential feature of XML, nor is it encouraged by the IEEE standard. Words in all tags are generally spelled out in full, except when they have a recognized abbreviation defined in the specification (e.g. "Movt" for "movement").

7. No specific units or system of units are implied in any of the tags that constitute the TMML vocabulary. For example, total delay may be

expressed in vehicle-seconds, vehicle-minutes or vehicle-hours, depending on the individual software product. It is essential that programmers understand the interpretation that each software product places on all data items.

To avoid setting rigid and prescriptive requirements, alternatives that encompass a broad range of current practice are offered. No software product is likely to recognize all of the data elements contained in the TMML specification. Each product that offers TMML connectivity will require a programmer interface document, which identifies the data elements that are recognized and any conditions that apply to their interpretation.

Description of TMML Classes

Each of the classes represented in Figure 5-1 provides a "container" for the data elements that describe the properties of the class. The classes that represent an intersection facility include the following:

1. The GENERAL Class, which contains elements that describe the file itself, as opposed to the traffic data within the file.

2. The AGENCY Class, which contains a collection of elements that describe the user of the program.

3. The INTERSECTION Class, which contains all of the sub-classes and data items required to represent each intersection. The attribute value must indicate the specific intersection to which the data elements apply.

4. The CONTROLLER Class, which contains all of the data elements that apply to the controller as a whole. This class is a child of the

INTERSECTION class. There is no attribute because there is only one controller per intersection.

5. The TIMINGPLAN Class, which contains all of the information required to describe one phase of the timing plan at an intersection. The applicable signal phase must be specified as an attribute.

6. The PHASECODES Class, which provides options to accommodate the signal phasing schemes used by the common signal analysis software. The representation of signal phasing differs widely among traffic modeling software.

7. The APPROACH Class, which establishes the specific approach to the intersection. The approach designation must be included as an attribute.

8. The LANEGROUP Class, which establishes a subset of a specific approach. A lane group is a set of one or more lanes carrying a homogeneous mixture of movements.

9. The MOVEMENT Class, which identifies a specific movement (left, through, right) as a child of the APPROACH class. The specific movement must be designated as an attribute.

LIST OF REFERENCES

1. Highway Capacity Manual, Special Report 209, Transportation Research Board, National Research Council, Washington, DC, 1997.

2. 1998 Level of Service Handbook, System Planning Office, Florida Department of Transportation, Tallahassee, FL, 1998.

3. Traffic Engineering Manual-Metric Version, Traffic Engineering Office, Florida Department of Transportation, Tallahassee, FL, 1996.

4. Traffic Control System Handbook, Federal Highway Administration, Washington, DC, 1985.

5. McShane, W. R. and R. P. Roess, Traffic Engineering, Prentice Hall, Englewood Cliffs, NJ, 1990.

6. ITT Industries, CORSIM User's Manual, version 4.3, Office of Safety and Traffic Operations R & D, Intelligent Systems and Technology Division (HSR-10), Mclean, VA, June, 1999.

7. Highway Capacity Manual, Special Report 209, Transportation Research Board, National Research Council, Washington, DC, 2000.

8. Webster, F. V., Traffic Signal Settings, Road Research Technical Paper No. 39, Road Research Laboratory, Her Majesty's Stationery Office, 1958. Reprinted with minor amendments, 1961.

9. Pignataro, L. J., Traffic Engineering: Theory and Practice, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1973.

10. Akçelik R., Rahmi, Besley and Mark, SIDRA 5, Australian Road Research Board, Transportation Research, Vermont South, Victoria, Australia, June, 1996.

11. Husch, D., Synchro 3.2 User Guide, Trafficware, Berkeley, CA, 1998.

12. Courage, K. G. and C. E. Wallace, TRANSYT-7F Users Guide, Federal Highway Administration, University of Florida, Gainesville, FL, December 1991.

13. Saito, M. and J. Fan, Artificial Neural Network – Based Heuristic Optimal Traffic Signal Timing, Computer-Aided Civil and Infrastructure Engineering, Volume 15, Issue 4, Malden, MA, 2000, pp 281-291.

14. Highway Capacity Manual, Special Report 209, Transportation Research Board, National Research Council, Washington DC, 1985.

15. Highway Capacity Manual, Special Report 209, Transportation Research Board, National Research Council, Washington, DC, 1994.

16. Akçelik, R. Signal Timing Calculation Methods for Vehicle-actuated and Fixed Time Signals, ARRB Transportation Research, TO WD95/020, Vermont South, Victoria, Australia, December, 1995.

17. Akçelik, R. and E. Chung, Calibration of Performance Models for Traditional Vehicle-actuated and Fixed Time Signals, ARRB Transportation Research, WD TO95/013, Vermont South, Victoria, Australia, December, 1995.

18. Courage, K. G., D. F. Fambro, R. Akçelik, P. S. Lin, M. Anwar and F. Viloria, Capacity Analysis of Traffic-Actuated Intersections, NCHRP Project 3-48 Final Report, University of Florida, Texas Transportation Institute and ARRB Transportation Research, National Cooperative Highway Research Program, Transportation Research Board, National Research Council, USA, December 1996.

19. Kell, J. H. and I. J. Fullerton, Manual of Traffic Signal Design, second edition, Institute of Transportation Engineers, Prentice Hall, Englewood Cliffs, NJ, 1991.

20. Lin, F. B., Optimal Timings Settings and Detector Lengths of Presence Mode Full-Actuated Control, Transportation Research Record 1010, Transportation Research Board, National Research Council, Washington DC, 1985, pp37-45.

21. Courage, K. G., J. Z. Luh and C. E. Wallace, Development of Guidelines for Implementing Computerized Timing Designs at Traffic Actuated Signals, 2 Volumes on Arterial System Implementation, Transportation Research Center, University of Florida, Gainesville, FL, 1989.

22. Orcutt, F. L., The Traffic Signal Book, Prentice Hall, Englewood Cliffs, NJ, 1993.

23. Bullen, A. G., R. N. Hummon, T. Bryer and R. Nekmat, EVIPAS: A Computer Model for the Optimal Design of a Vehicle-Actuated Traffic Signal, Transportation Research Record 1114, Transportation Research Board, National Research Council, Washington DC, 1987, pp 103-110.

24. EVIPAS 1.0 Users Guide, Viggen Corporation, Sterling, VA, 1993.

25. Husch, D., The Percentile Delay Method, unpublished, Berkeley, Berkeley, CA, August, 1996.

26.   Husch, D., Comparing Signal Delay Method, unpublished, Berkeley, CA, May, 1997.

27.   Husch, D., Comparing Capacity Analysis, unpublished, Berkeley, CA, January, 1998.

28.   Husch, D., SYNCHRO User Guide, Trafficware, Berkeley, CA, February, 1998.

29.   Chundury, S. and B. Wolshon, Evaluation of the CORSIM Car-flowing Model Using GPS Field Data, Transportation Research Record 1710, Transportation Research Board, National Research Council, Washington, DC, 2000, pp 114-121.

30.   Courage, K. G. and C. E. Wallace, The M|O|S|T Reference Manual, Federal Highway Administration, University of Florida, Gainesville, FL, December, 1991.

31.   Fu, L., Generic Search, Neural Networks in Computer Intelligence, McGraw-Hill, Inc., New York, NY, 1994, pp239-240.

32.   Gen, M. and R. Cheng, Genetic Algorithms and Engineering Design, Ashikaga Institute of Technology, Wesley Publishing Company, Inc., New York, NY, 1989.

33.   Goldberg, D. E., Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley Publishing Company, Inc., New York, NY, 1989.

34.   Man, K. F., K. S. Tang, S. Kwong and W. A. Halang, Genetic Algorithms for Control and Signal Processing, Springer-Verlag, London, UK, 1997.

35.   Abu-Lebdeh, G. and R. F. Benekohal, Development of a Traffic Control and Queue Management Procedure for Oversaturated Arterials, Transportation Research Record 1603, Transportation Research Board, National Research Council, Washington, DC, January, 1997, pp 119-127.

36.   Hadi, M. A. and C. E. Wallace, Hybrid Genetic Algorithms to Optimize Signal Phasing and Timing, Transportation Research Record 1421, Transportation Research Board, National Research Council, Washington, DC, January, 1997.

37.   Park, B., C. J. Messer and T. Urbanik II, Traffic Optimization Program for Oversaturated Conditions: A Genetic Algorithm Approach, Transportation Research Record 1683, Transportation Research Board, National Research Council, Washington, DC, 1999, pp 133-142.

38.   Ginsberg, M., Essentials of Artificial Intelligence, Morgan Kaufmann Publishers, San Francisco, CA, 1993.

39.   Hillier, F. S. and G. J. Lieberman, Introduction to Operations Research, Fifth Edition, McGraw-Hill, Inc., New York, NY, 1990.

40. Bazaraa, M. S., J. J. Jarvis and H. D. Sherali, Linear Programming and Network Flows, John Wiley & Sons, New York, NY, 1990.

41. Leonard II, J. D., CORTOOLS Automation Server, unpublished, Georgia Institute of Technology, Atlanta, GA, 1998.

42. Courage, K. G., Signal Operations Analysis Package for 2000 (SOAP2K) Functional Requirements, Working Report, University of Florida, Gainesville, FL, 1999.

43. Li, M. and A.C. Gan, Signal Timing Optimization for Oversaturated Networks Using TRANSYT7-F, Transportation Research Record 1683, Transportation Research Board, National Research Council, Washington DC, 1999, pp 118-126.

44. Kim, J.T., Equilibrium Models for Shared Lane Permitted Left Turns, Master's Report, Department of Civil Engineering, University of Florida, Gainesville, FL, 1997.

45. Mood, A. M., F. A. Graybill and D. C. Boes, Introduction to the Theory of Statistics, McGraw-Hill, Inc., New York, 1974.

46. Wu, J., Evaluation and Enhancement of Opposed Left-turn Flow Analysis Methodology, Ph.D. Dissertation, Department of Civil Engineering, University of Florida, Gainesville, FL, 1995.

47. Courage, K. G., TMML Model Markup Language, Transportation Research Center, University of Florida, Gainesville, FL, July, 2000.

48. Kim, J. T. and K. G. Courage, Development and Application of a Traffic Model Data Structure Based on XML, Transportation Research Center, University of Florida, Working Paper for WCTR (C43411), Gainesville, FL, 2001.

49. Bosak, J. and T. Bray, XML and the Second-Generation Web, Scientific American, May, 1999.

50. Harold, E. R., XML Bible, IDG Books, Foster City, CA, 1999.

51. Laurent, S., XML: A Primer, Second Edition, MIS Press, M & T Books, Foster City, CA, 1999.

52. Simpson, J. E., Just XML, Prentice Hall PTR, Upper Saddle River, NJ, 1999.

53. Draft Standard for Data Dictionaries for Intelligent Transportation Systems-Part 1 Functional Area Data Dictionaries, Intelligent Transportation Systems Standards Coordinating Committee (SCC 32), IEEE P1489/D0.1.2, New York, NY, February, 1999.

54.   ITT Systems, Background Material for Presentation of ARADS Status to the TSDTF, Working Paper, Colorado Springs, CO, March, 1999.

BIOGRAPHICAL SKETCH

Jin-Tae Kim was born in Seoul, Korea, R.O.K., on November 27, 1969. In February 1992, he received his Bachelor of Science degree from the Department of Transportation at Hanyang University. During his study, he received several honor student scholarships from the university. In 1991, he was certified as a top grade transportation engineer by passing the tests given by the Korean Manpower Agency. In January 1992, he enrolled in the Graduate School of the University of Seoul as a transportation-engineering major in the Department of Urban Engineering by being awarded an honor student scholarship from the university. He accepted a position as a transportation engineer in April 1992 at the Research Institute of Engineering and Technology at Hanyang University. He worked for the projects issued by the Korean Freeway Corporation. In January 1993, he was commissioned as a discipline trainer of cadets for the mandatory military service at the Korean Third Military Academy.

In January 1996, he enrolled at the University of Florida as a graduate student in the Department of Civil Engineering with a major emphasis in transportation engineering. During his study for the master's degree, he worked for the Florida Technology Transfer Center as a graduate assistant. In August 1997, he received his Master of Engineering degree from the University of Florida.

He continued his study at the University of Florida to pursue the degree of Doctor of Philosophy in the Department of Civil Engineering with a major emphasis in

transportation engineering. During his study, he was honored to participate as a graduate research assistant in the project of the Signal Operation Analysis Package 2000 (SOAP2K) development in the Transportation Research Center at the University of Florida. He was also honored to be involved in various projects issued by the Florida Department of Transportation and won first place for the best student paper from District 10 of the Institute of Transportation Engineers in November 1998. In August 2001, he graduated with the degree of Doctor of Philosophy.
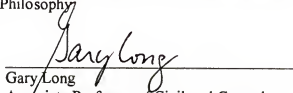
He married Hee-Jae Na while he studied at the University of Florida. He is a student member of the Institute of Transportation Engineers, the American Society of Civil Engineers, the Korean Transportation Association in America and the Korean Society of Transportation.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Kenneth G. Courage, Chair
Professor of Civil and Coastal Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Gary Long
Associate Professor of Civil and Coastal Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
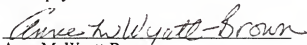
Scott Washburn
Assistant Professor of Civil and Coastal Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Joseph P. Geunes
Assistant Professor of Industrial and System Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Anne M. Wyatt-Brown
Associate Professor of Linguistics

This dissertation was submitted to the Graduate Faculty of the College of Engineering and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December 2001

Pramod P. Khargonekar
Dean, College of Engineering

Winfred M. Phillips
Dean, Graduate School